The Web of Questions

Inquisitive Decision Theory and the Bounds of Rationality

by

Daniel Hoek

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

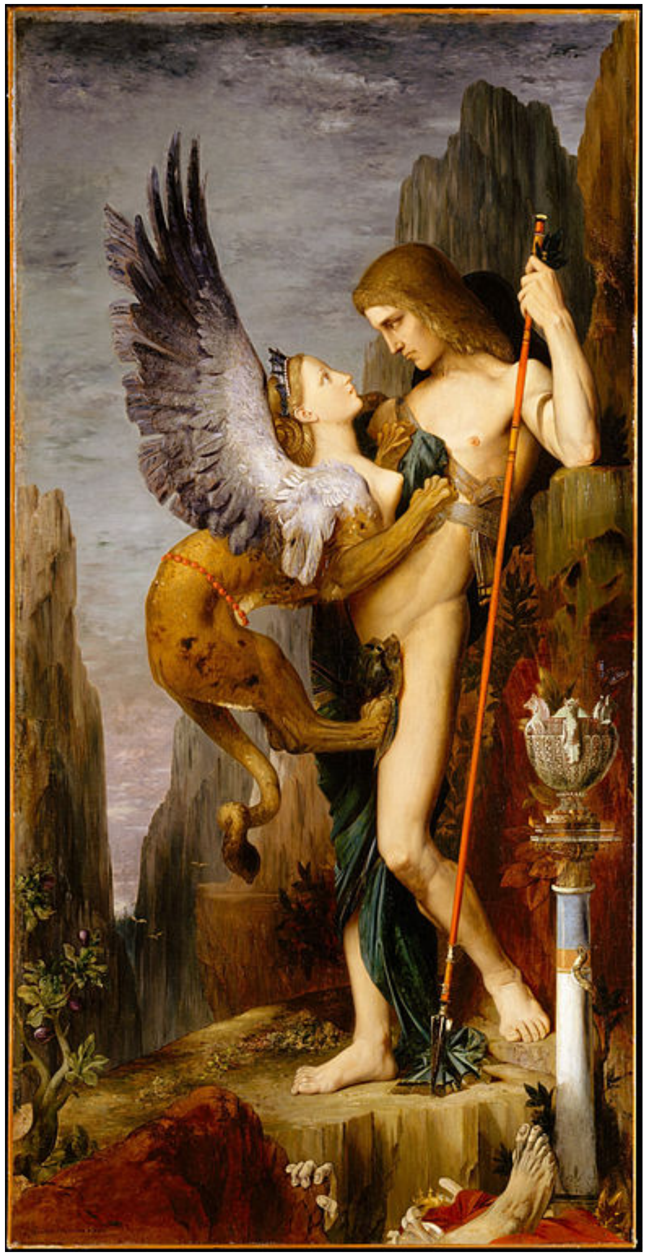
Department of Philosophy

New York University

September 2019

____

Cian Dorr, advisor



Oedipus Interrogated by the Sphinx

Gustave Moreau, 1864, Metropolitan Museum of Art

Acknowledgments

My greatest debt is to my advisor Cian Dorr, who took a keen interest in this project from the moment I told him about it. Without his insistence to keep going when the going was tough, these ideas might never have come to fruition at all. And I am certain that without his insights, knowledge, and advice, the product would have been incomparably worse. His contributions to the content of this dissertation are substantial.

For the long conversations, criticism and for detailed comments on drafts of (parts of) this material, I am indebted to my other committee members, Chris Barker and David Chalmers, and also to Justin Bledin, Paul Egré, Adam Elga, Jane Friedman, Harvey Lederman and Salvador Mascarenhas. In addition, I have benefited from exchanges with Martín Abreu Zavaleta, Maria Aloni, Amir Anvari, Bob Beddor, Johan van Benthem, Kyle Blumberg, Christa Carbo, Sam Carter, Ivano Ciardelli, Milica Deniç, Matti Eklund, Diego Feinmann, Hartry Field, Kit Fine, Carolina Flores, Nils Franzén, Iliana Gioulatou, Jeremy Goodman, Ian Grubb, Ben Holguín, Zoë Johnson King, Dirk Kindermann, Arden Koehler, Manuel Križ, Uriah Kriegel, Jeremy Kuhn, Adam Lerner, Slawa Loev, Adam Lovett, Jordan MacKenzie, Rosie McHokey, Tricia Magalotti, Mora Maldonado, Annette Martin, Berislav Marušić, Matt Moss, Morgan Moyer, Kevin Munger, Michael Murez, Jim Pryor, Agustín Rayo, François Recanati, Daniel Rothschild, Chris Scambler, Stephen Schiffer, Ginger Schultheis, Anthia Solaki, Benjamin Spector, Andrés Soria Ruiz, Bob Stalnaker, David Storrs-Fox, Trevor Teitel, Silvia De Toffoli, Daniel Viehoff, Jerome Wakefield, Daniel Waxman, Steve Yablo and Seth Yalcin. For their insightful questions, I thank audiences at New York University, the Institut Jean Nicod, the University of Graz, the University of Amsterdam (ILLC), University College London, Brandeis University, the University of Pittsburgh, the New York Philosophy of Language Workshop, the National University of Singapore, Bilkent University, Virginia Tech and Uppsala University.

For all the intellectual support and excitement they supplied over the past six years, I want to thank everyone at the NYU philosophy department, as well as my other friends and colleagues in New York. I want to extend a special word of thanks to the speakers, regulars and organisers at my beloved New York Philosophy of Language Workshop. Much of this dissertation was written during repeat visits to the Institut Jean Nicod, which is an extraordinary group of philosophers, linguists and cognitive scientists at the École Normale Supérieure in Paris. I am extremely grateful to the faculty and students of the IJN for letting me be a part of it.

My fascination with issues around the problem of logical omniscience began while I was an undergraduate. For this I am indebted to Brian King, who introduced me to the mysteries of mathematical knowledge and ignorance; to Alex Anthony, who got me hooked on Stalnaker; to Bob Stalnaker, whose writings showed me the awesome depth of these issues; and to Steve Yablo, whose Locke lectures revealed a glimpse of a solution.

Abstract

Something important is missing from the standard account of the connection between belief and action. The way a given belief should be expected to manifest itself in action is not a function of its informational content, but also depends systematically on the *question* that the belief answers. This dissertation articulates that dependence with a simple new theory of belief-guided action, explaining a range of ordinary patterns of behaviour that cannot be accounted for given the standard account of belief, desire and action. The appeal to questions is especially fruitful when it comes to explaining behaviour that displays some inconsistency, or which is less than ideally rational. I call this new account of belief and action *inquisitive decision theory*.

Besides providing a new model for less than ideally rational behaviour, the inquisitive account of belief also suggests new ways of thinking about deductive reasoning and deliberation, and throws new light on certain long-standing issues in doxastic logic. It brings together a converging set of recent insights about the role of questions in cognition stemming from epistemology, the philosophy of language, the metaphysics of propositions, linguistic semantics, formal pragmatics and psychology. In addition, it builds on work in decision theory, computer science, behavioural economics and the philosophy of mathematics.

Table of Contents

List of Figures

List of Tables

List of Symbols

The numbers in brackets refer to the relevant definitions.

w, v, u	Worlds
ϕ, ψ, χ	Clauses/Sentences
α, β, γ	Agents
p, q, r, …	Intensional propositions (sets of possible worlds)
Q, R, S, …	Questions (1.5)
A, B, C, …	Answers (1.5)
A^{Q}, B^{R}, C^{S}, …	Quizpositions (1.6)
$\mathsf{AB}^{\mathsf{QR}}$, $\mathsf{CD}^{\mathsf{ST}}$, …	Conjunctive quizpositions (2.7)
A/R	The maximal R-part of A (2.9)
$\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, …	Options (1.7)
Γ, Δ	Decision problems (1.7)
$\mathscr{W}$	Logical space (the set of all possible worlds)
$\mathscr{P}(\mathscr{W})$	The space of intensional propositions (subsets of $\mathscr{W}$)
$\mathscr{D}$	A domain of questions closed under parthood (2.11, 3.6)
$\mathscr{Q}(\mathscr{D})$	The space of quizpositions about questions in $\mathscr{D}$

S, **T**	Sets of propositions/quizpositions
I, **J**	Information states (2.1, 2.10)
$\mathbf{B}_\alpha$	Belief state (2.2, 2.12)
$\mathbf{CB}_\alpha$, $\mathbf{QB}_\alpha$	Classical/Inquisitive belief state (4.6/4.13)
Pr	Probability (3.1, 3.6)
$\mathbf{Cr}_\alpha$	Credence state (3.3, 3.8)
$\mathcal{E}_{\mathbf{Pr}}$, $\mathcal{E}_{\mathbf{Cr}}$	Expected Value (3.2, 3.7) / $\mathcal{E}$-value (4.26)
$\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$	Agential states (4.2)
$\succ_\alpha$	Preference (4.3)
$\succ_\alpha^{\mathrm{Q}}$	Q-Preference (4.4)
$>$	Strict dominance (4.5)
$>_{\mathrm{p}}$	Strict p-dominance (1.9, 4.5)
$>_{\mathrm{p}}^{x}$	Strict x-p-governance (4.9)

Introduction

Imagine you find yourself in the middle of a cold, inhospitable forest at dusk, bereft of supplies and surrounded by disheartening animal noises. You come to a crossroads and have to choose a path. Hungry eyes are tracing you, and you face an almost palpable question: *How do I get out of here?* Questions await us at all the crossroads of life, both literal and metaphorical, even if they are usually less consequential. The choice of how many eggs to get at the supermarket raises the question *How many eggs go into a spaghetti carbonara for four?* Plotting your next chess move, you face the question *How do I put my opponent on the defensive?* In the flower shop, you wonder *What is his favourite colour?* And so on: whenever you make a choice, you face a question.

And what you decide to do normally depends on your answer to the question raised. Take the supermarket situation. If you reckon you need five eggs for your carbonara, you will buy half a dozen. If you think you need eight, you get a dozen. If you are unsure, maybe you still get a dozen just to be on the safe side. Thus the decision you make is guided by your answer to the question that the choice confronted you with. If you know the right answer to this question, you will generally make the right choice, while wrong answers lead to bad choices. If you are faced with a question that you have no answer to at all, then you are likely also unsure what to do.

In this dissertation, I develop this question-centric or *inquisitive* way of thinking about belief-guided action in a systematic way. The idea of connecting choices to questions is so natural and intuitive that it may appear innocuous. But it actually yields a substantial departure from the received view, making sense of a range of otherwise puzzling psychological phenomena, like the distinction between recognition and recall. Most importantly, the inquisitive picture suggests a systematic account of the behaviour of agents who fail to see some of the consequences of their beliefs, and agents with inconsistent beliefs.

The inquisitive theory I propose is modelled on the traditional, *classical* account of belief-guided action. This account has its fullest, most influential articulation in standard decision theory, interpreted the way economists and psychologists do: as an attempt to explain and predict behaviour in terms of an agents' beliefs and desires.[1] On that descriptive interpretation, decision theory is already highly controversial in exactly the sort of way that motivates the inquisitive turn. According to its critics, the strong rationality assumptions baked into the classical theory are so unrealistic that the theory cannot be trusted to yield accurate predictions about ordinary people. Celebrated exponents of that critique include Daniel Kahneman (2012) and Richard Thaler (Thaler and Sunstein 2008). Their groundbreaking work led to a call for less idealised,

[1] For the record, inquisitive decision theory *can* also be interpreted normatively, as an account of what we rationally ought to do given our (possibly inconsistent) beliefs and desires. I think this normative interpretation is both interesting and well-motivated, but it will not be my focus here.

more realistic accounts of decision making. The inquisitive theory of belief-guided action is one answer to that call.

You can think of both the inquisitive and the classical theory as consisting of three nested, interrelated components. Each of the first three chapters concerns one of these components. The core component of each theory is a view about the content of a belief, and about the way an individual full belief affects an agent's choices: these views are the topic of Chapter 1. The second component is a distinctive view of the way individual beliefs come together in belief *states*, to be described in Chapter 2. The third and final component is the account of doxastic uncertainty and *credences* or partial beliefs, to be covered in Chapter 3. With this third component in place, my proposal takes the form of a novel, *inquisitive* decision theory. Along the way, we build an increasingly detailed understanding of the way our cognitive limitations manifest themselves in the inquisitive theory, and in Chapter 2 I introduce and examine a simple inquisitive model for deductive reasoning.

In Chapter 4, I turn to some formal results that underpin the theory articulated in the first three chapters. I will provide a unified formal framework for thinking about behavioural dispositions generally, and use this framework to prove representation theorems for both classical and inquisitive decision theory. These formal results give us a precise way of seeing just how much less idealised the inquisitive decision theory is, allowing a direct comparison of the rationality assumptions that go into classical and inquisitive decision theory.

The decision-theoretical issues that are the central motivation in Chapter 1-4 intersect with a longstanding issue in doxastic logic known as the *problem of logical omniscience*. Roughly speaking, this is the problem of constructing a tractable and realistic formal model of the beliefs and credences of agents who do not know every consequence of their beliefs, and whose beliefs may be inconsistent. In addition to its decision-theoretical manifestation, the problem of logical omniscience has many other faces: it is a complex, many-faceted issue at the intersection of psychology, economics, philosophy and linguistics. I cannot attempt to do justice to every aspect of the problem explicitly within this dissertation. However, I do think the inquisitive theory of belief set out below has the potential to cast new light on many of its manifestations. In Chapter 5, I illustrate this with an excursion beyond decision theory.

Chapter 5 concerns the problem of *mathematical* omniscience. According to the classical view of belief states, an agent's beliefs are closed under entailment (necessitation). In particular, that means everyone believes every necessary truth, which includes every mathematical truth. But of course we do not know every mathematical truth, and this difficulty gives rise to the problem of *explicit* mathematical omniscience. The other half of the problem is that classical agents also cannot manifest any *implicit* mathematical ignorance. For instance, if a classical agent knows the diameter of Martha's perfectly circular yard is 12 feet, it would follow that they also know the circumference of Martha's yard to arbitrarily many decimal places.

The inquisitive account of belief and credence provides a more promising basis for an account of mathematical belief. It does not face an implicit problem of mathematical omniscience at all, and I argue that it brings us an important step closer to addressing the explicit problem as well. Some of the considerations about the problem of mathematical omniscience are also relevant to

Frege's Puzzle, which can be regarded as another manifestation of the problem of logical omniscience. In the final section of Chapter 5, I draw those connections.

Apart from a few stray remarks, what is missing from the dissertation is a treatment of the semantics of belief *reports* — the sentences we use to describe and attribute beliefs. This would be a surprising omission in any philosophy dissertation about belief, but especially coming from someone who specialises in the philosophy of language. My excuse for this hiatus is that it reflects what I have come to think is the proper order of inquiry here. The problem of logical omniscience as it arises in semantics and doxastic logic can formally be addressed using a wide variety of fine-grained conceptions of belief content. But such formal solutions do not touch the root of the problem unless they explain what it is about our doxastic mental states that makes it the case that belief contents are fine-grained in this or that particular way. As Stalnaker put it, "One needs an account of what states of belief, desire and intention *are* that explains how the fine-grained structure of some notion of proposition contributes to distinguishing between different states of belief, desire or intention." (Stalnaker 1999b, p. 27)

Below I provide such an explanation for the kind of inquisitive propositional structure I favour. The account certainly suggests a moral for the semantics of belief reports, and it provides the conceptual groundwork for a new semantics. But I will remain agnostic about the details of how this moral is to be implemented. Should we to appeal to alternatives semantics, to questions under discussion, or inquisitive semantics? What is the relation to embedded questions in knowledge reports? How ambiguous are belief and knowledge attributions? What is the role of focus? These are all great questions. But we will face them another day.

Chapter 1.
The Inquisitive Turn

In this chapter, I aim to establish the basis of the classical and inquisitive pictures of belief-guided action, and to provide some initial motivation for the move to the latter. We will see that there are certain ordinary, everyday patterns of behaviour that the classical picture cannot make sense of, but which have a perfectly natural and straightforward explanation on the inquisitive picture.

More specifically, this chapter investigates the ways in which a given outright belief can be expected to guide an agent's actions. I will present and contrast the classical and inquisitive view of the contents of a belief, and the classical and inquisitive view of the way a belief with a certain particular content manifests itself in action. As we will see in subsequent chapters, all other aspects of the inquisitive theory of belief-guided behaviour follow on very naturally once you make this fundamental step from the classical to the inquisitive conception of individual beliefs. Thus the entire dissertation is essentially a drawn-out exploration of the ramifications of the inquisitive turn I argue for in this chapter.

Here is a concise statement of the view of individual beliefs at the heart of the classical picture:

> **Individual Beliefs (Classical):** A belief is the possession of a piece of information about the world, and manifests itself in behaviour as a general disposition to act on that information. (1.1)

Different authors have advocated different versions of this view, so I have intentionally left the initial statement a bit ambiguous. In particular, a "piece of information" could be a sentence, or a Russellian proposition, or a set of possible worlds. In one form or other, (1.1) has been endorsed, at least as a *ceteris paribus* approximation, by theorists of belief of all persuasions: from Fodor to Friston, from Hempel to Hintikka and from Savage to Schwitzgebel. (I provide a brief overview below.)

I will argue that, if we really want to address the fundamental problems of the classical picture, even this most basic, uncontroversial part stands in need of qualification. On the alternative view I propose, the behavioural manifestation of a given belief does not just depend on the information it carries, but also on the question it answers:

> **Individual Beliefs (Inquisitive):** A belief is the possession of an answer to a specific question, and manifests itself in behaviour as a disposition to act on that answer when faced with that question. (1.2)

As will become clear over the next three chapters, this seemingly modest adaptation makes a big difference, allowing us to account for a range of ordinary behaviours that cannot be classically explained.

Here is a first illustration (inspired by Elga and Rayo 2019, §2):

> ROMEO RECALL: Juliet comes home to find a note saying "*Somebody called for you — didn't catch a name but he sounded upset.*" There is a phone number below it, but the beginning is smudged out and Juliet can only read the final digits "6300". She instantly recognises Romeo's number, and she decides to go see him. When no-one answers the door, she rushes to a nearby phone booth in order to call him back. She dials 2-1-2-5-2-9-… only to realise she does not remember those last four digits.

The classical view has trouble accounting for Juliet's behaviour here. First she acts on the information that *Romeo's number ends in -6300*, later she does not. So does she believe it or not? If she has the belief, we have no classical explanation for why she failed to act on it in the phone booth. If she lacks it, then why did she go to Romeo's house? Either way, we have no complete classical account of her actions.

With respect to this example, the advantage of the inquisitive picture lies in the fact that it makes a distinction between two potential beliefs that the classical view identifies. One is the belief that ***Romeo****'s number ends in -6300*, in answer to the question *Who has a number ending in -6300?* The other is the belief that *Romeo's number ends in* ***-6300***, in answer to the question *What is Romeo's number?* These two belief contents are informationally equivalent, but inquisitively distinct. Thus the inquisitive view allows us to say that Juliet has the former belief but lacks the latter. The note confronts Juliet with the question *Who has a number ending in -6300?* Her response is guided by her answer to that question: *Romeo*. In the phone booth, Juliet is intuitively confronted with a different question, *What is Romeo's number?* It is because she has no answer to the latter question that she does not act on this information on that occasion. In this way, the inquisitive view accounts for the contrast between Juliet's behaviour in these two contexts on the basis of the fact that she is confronted with a different question on each occasion.

A defender of the classical view might insist that by the time Juliet arrives at the phone booth, she has lost the information she had before: she believed at first that *Romeo's number ends on -6300*, but by the time she made it to the phone booth, she must have forgotten. However, this kind of response is insufficiently general to be compelling. For if Juliet had really lost this information, she would not be able to *recognise* the phone number at this point either. But imagine Juliet were to spot the digits "6300" scribbled down on the wall of the phone booth. Probably, she would breathe a sigh of relief and dial the number. This shows that she has in fact retained the ability to recognise the number. Furthermore, the forgetful explanation assumes that when Juliet came home, she *did* have an ability to reproduce the last digits of Romeo's number. But that need not be the case. The bottom line is that classical picture is fundamentally ill-equipped to capture the familiar and psychologically well-established distinction between the ability to recognise a phone number and the ability to recall it. By contrast, that distinction comes naturally on the inquisitive picture.

The thesis that cognitive content is not purely informational but in some way tied to questions has ample independent motivation. It has been defended on a wide variety of epistemological, linguistic and psychological grounds by Dretske (1970), Schaffer (2004), Yalcin (2008, 2011, 2018), Egré and Bonnay (2012), Blaauw (2013), Koralus and Mascarenhas (2013, 2018), Yablo (2014), Fritz and Lederman (2015), Pérez Carballo (2016), Friedman (2017), Simons, Beaver, Tonhauser and Roberts (2017), Bledin and Rawlins (2018) and Holguín (2019). In particular, Seth Yalcin's and Alejandro Pérez Carballo's inquisitive models of belief anticipate certain aspects of my account, and are likewise motivated by concerns related to logical omniscience.

Robert Stalnaker's (1991) proposal for a question-and-answer model of belief is a more ambiguous precedent. He describes the idea only to reject it thus: "even if we had a satisfactory account of accessibility for the question-and-answer model, it would not be clear how to generalize it to an account of knowledge and belief in terms of capacities and dispositions to use information (or misinformation) to guide not just ones question-answering behaviour, but one's rational actions generally. *For we want an account of knowledge and belief, not just for expert systems and people who staff information booths, but for all kinds of agents*." (Stalnaker 1991, p. 253-4, my emphasis.) In this dissertation I seek to answer Stalnaker's challenge, providing an account that has the generality he demands here. To my knowledge, it is the first attempt at a systematic and general theory of the role of questions in practical deliberation and decision making.

As I said, the classical view (1.1) I am looking to adjust is widespread, and comes in many forms. For Hartry Field (1978), the pieces of information are declarative sentences in natural language, and belief involves the storage of such sentences in memory so that they can be used in deliberation. Jerry Fodor (1978, 1980) held that beliefs are sentences in an internal language whose tokens causally affect the agent's actions. Fred Dretske (1988, ch. 4) said the propositions expressed by those internal representations explain the behaviour, not the representations themselves. Daniel Dennett (1971, 1975) agreed that propositions do the explaining, and claimed the internal representations could therefore be disposed of. And "propositions" could be many things: syntactically structured entities (Carnap 1947, Cresswell 1985, King 2007), map-like entities (Camp 2007), sets of possible worlds (Hintikka 1962, Stalnaker 1984), sets of (im)possible worlds (Barcan Marcus 1983, Smets and Solaki 2018), or sets of world-plan pairs (Gibbard 2003).

Besides all these different conceptions of information, there is a further dimension of variation worth highlighting, namely the way belief interacts with other attitudes to arrive at a decision. Standard belief/desire psychology has it that beliefs combine with desires, goals, plans or utilities to arrive at a decision. But on Alan Gibbard's Aristotelean view, agents have beliefs about "what to do," making decisions on the basis of their beliefs alone, without the admixture of a separate desire attitude (Gibbard 2003). According to Michael Bratman (1984, 1999), the expression of beliefs and desires through action is mediated: intentions and plans are the direct guides of action, and beliefs and desires affect action indirectly, by constraining the formation of those intentions and plans. Other endorsements of the classical view (1.1) in one form or other can be found in Peirce 1878, Ramsey 1926, Hempel 1962, Savage 1972, Millikan 1986, Schwitzgebel 2002 and Friston and Mathys 2017.

I do not pretend to be giving a blanket refutation of all these multifarious theories of belief. Rather, I am offering a positive suggestion: I think that any account of belief stands to gain from acknowledging the role of questions, and one aim of this chapter is to motivate theorists of belief of all stripes to join the inquisitive turn. Most of those just cited endorse the classical view (1.1) only as a rough, *ceteris paribus* generalisation about the connection between belief and action. The inquisitive view (1.2) is intended as an improvement on that coarse approximation. What is more, it is an improvement that can achieved without going beyond the intentional level. That is to say, (1.2) still accounts for a belief's role in behaviour purely in terms of its content, rather than resorting to lower-level types of explanation in terms of the computational or neurological realisation of the belief.

In sections §1.1-2 below, I will examine the motivation for the inquisitive view of individual beliefs in greater detail. The remainder of the chapter exhibits a formal and precise way to articulate the inquisitive view (1.2). §1.3 gives an account of inquisitive belief contents, saying just what is meant by "an answer to a question". §1.4 says what it is for an agent to *face a question*, and for a particular choice to *confront* an agent with a question. §1.5 completes the main task of the chapter, discussing what, exactly, it takes for an agent to *act on* an answer to a question. §1.6-7 examine the decision-theoretical framework set up in §1.4 a bit more carefully, bringing out some issues that the initial presentation glosses over.

1.1 Recognition and Recall

In the ROMEO RECALL example we just examined, Juliette succeeds in recognising the last digits of Romeo's phone number but is unable to reproduce them. With phone numbers that is often the way: recognition is somewhat easier than recall. But as the following story shows, that is not always the case (based on Elga and Rayo 2019, Powers 1978):

> TRIVIAL TROUBLE: Travelling abroad, Tom is accosted by a fearful sphinx. "Don't worry yourself darling," she says, pinning him gently but firmly to the ground, "I've eaten. But I will make you wealthy if you solve me this riddle: name an English word that ends in the letters -MT." Tom racks his brain, but in the end

> admits defeat. Turning to leave, the sphinx remarks: "You're not the brightest bulb, are you now? I could have done this in my sleep!" That night, Tom writes a letter home: "I never dreamt that sphinxes were real!" Then he reads it back incredulously: "D-R-E-A-M-T".

Here Tom fails to *recognise* the word "dreamt" based on its last two letters, although he later has no trouble *reproducing* those same letters. As with Juliet, when we try to analyse Tom's behaviour classically, we quickly run into trouble. The issue this time is whether or not Tom has the information that *the word "dreamt" (or /drɛmt/) is spelt D-R-E-A-M-T*. On the one hand, Tom dramatically fails to act on this information during his encounter with the sphinx: else he would have answered correctly and walked away with wealth untold. So to explain Tom's actions at that point classically, we must say he lacks this information. But then how is he able to spell the word correctly later on? How can Tom be acting on the information then, unless he had it all along? The upshot is that, whether we say that Tom has this information or not, we cannot explain both actions classically. There is no satisfying classical explanation for the contrast between Tom's behaviour when quizzed by the sphinx and when writing his letter.

The inquisitive picture, by contrast, yields a straightforward account of this case: the sphinx confronts Tom with the question *Which words end on -MT*; but when writing his letter, he is intuitively confronted with a different question: *How to spell "dreamt"*. Tom knows *"dreamt" ends in **-MT*** in response to the latter question; but that is no guarantee that Tom also has the informationally equivalent answer to the former question — and as his behaviour shows, as a matter of fact he does not.

At the end of TRIVIAL TROUBLE, Tom reads back his letter with the sphinx's question in mind, and experiences what he would describe as an epiphany. It is clear what to say about this on the inquisitive view: Tom has found the answer he was missing before, that ***"dreamt"*** *is a word ending on -MT*. On the classical picture, it is hard to say what changed for Tom. The only classical way to discover something new is to obtain a new piece of information. But what could it be? It cannot be the information that *"dreamt" ends in -MT*: Tom had it ever since he learned to spell. So the classical theorist looks to be at a loss to say what Tom's discovery consists in, or what explains his newfound ability to answer a question he could not answer before.

An analogous point applies ROMEO RECALL. Suppose we add a happy ending to the story: Juliet thought to bring the note with her! She takes it out of her pocket in the phone booth, and dials the digits she could not recall. But what does Juliet expect to learn from reading the note, exactly? Once again, only the inquisitive view has a satisfying answer. Thus both case studies demonstrate a further important advantage of the inquisitive view: it allows for the possibility of forming new beliefs without acquiring new information, something that should not be possible on the classical view. In Chapter 2, we will see how this feature of the inquisitive picture can be used to help us understand the nature of deductive reasoning.

The potential for more examples in the spirit ROMEO RECALL and TRIVIAL TROUBLE is boundless. For inspiration, consult the literature on memory retrieval or your nearest crossword. Song lyrics are also a nice source of examples. You might know the lyrics to a song, but that does not mean you will recognise the song based on a lyric, and vice versa. The more you look, the more you realise that our lives are designed around cognitive asymmetries that the classical picture has no accounting for.

1.2 Cognitive Failures and Classical Coherence

Both in the case of ROMEO RECALL and TRIVIAL TROUBLE, defenders of the classical picture may seek to explain away what seem like false predictions of their account by chalking them up to cognitive failures of some sort. Deep down, Tom *had* the answer to the sphinx's riddle, but due to some fluke or snag he was unable to use it. It was the nerves, maybe; or maybe he momentarily forgot. Whatever it was, some cognitive mishap prevented Tom's normal doxastic dispositions from displaying themselves on this occasions.

In response, let me begin by noting that Tom would have been equally floored by the question *Which words end on -MT* in the context of a crossword: the nerves are not really essential to the case. But more importantly, these attempts to explain Tom's actions in terms of cognitive failure seem to get other facts about his behaviour wrong. Firstly, if Tom's access to the information that *"dreamt" is spelt D-R-E-A-M-T* really was temporarily cut off, then Tom should have been equally stumped if the sphinx had asked him *How do you spell the word "dreamt"*. But in fact he could probably field *that* question easily. Secondly, these explanations imply that under less stressful circumstances, Tom *does* have the ability to reproduce the word "dreamt" based only on the letters -MT. But if Tom is like most English speakers, that is simply not true. Master scrabblers and crossword mavens aside, most people have great difficulty recognising words based on just a few letters. As it happens, this particular cognitive limitation is extremely well-documented, since psychologists frequently use word completion as a memory test: see for instance Phillips 1966, Nelson and McEvoy 1984, Olofsson and Nyberg 1995.

The cognitive failure response does not address the fundamental issue, which is that according to the classical picture, the ability to *reproduce* the final letters of "dreamt", if it is to be explained in terms of beliefs at all, should always come and go together with the ability to *recognise* the word on the basis of those letters. For on the classical picture, both abilities are different manifestations of the same belief: namely the belief that *"dreamt" ends in -MT*. But as a matter of psychological fact, these two cognitive abilities are basically independent. This phenomenon is far too robust and systematic to be written up to unexplained "slips of the mind". We need an account that does not lump these distinct cognitive abilities together, but instead associates them with different beliefs. The inquisitive picture does just that. It distinguishes two different ways of having the information that "dreamt" ends in -MT. You can have it in answer to the question *How to spell "dreamt"*: that belief gives you the ability to reproduce at least those two letters. Or you can have it in answer to the question *Which words end in -MT*. This latter belief is far less useful and far less common: having it allows you to recognise the word "dreamt" on the basis of its last two letters.

The classical view of individual beliefs (1.1) assumes the following coherence principle:

> **Classical Coherence**: If on some occasion, an agent acts on a piece information because of their beliefs, then that agent is in general disposed to act on that information. (1.3)

In other words, it is all or nothing: either we have no belief-based explanation of the fact that an agent acted on some information, or we are on the hook for a general disposition. That is what got the classical view into trouble in the examples above: we cannot give a classical doxastic explanation for Tom's spelling of the letter, or for Juliet's response to the note, without attributing to these agents a broad disposition that they simply do not possess.

A common response to cases like ROMEO RECALL and TRIVIAL TROUBLE is to jettison classical coherence entirely. The temptation is to allow pervasive failures of the disposition to act on the information you have. This makes for something like the following view: belief is the possession of a piece of information, manifesting itself in a disposition to act on that information *occasionally*. According to Davidson 1976, there is no systematic accounting for the exceptions: glitches galore! One can also try a more systematic approach. Perhaps we act on the information only when the belief in question is "active". Only when we think of the belief. Only when the belief makes it into your mental decision-making module. Only when the agent is paying attention to the right things. Only when the decision is in a certain practical domain. One sophisticated version of such an account is Elga and Rayo's (2019) fragmented decision theory, examined in §3.3 below. Another approach stems from the awareness literature in computer science and renders (explicit) belief relative to what the agent's state of attention (Fagin and Halpern 1988, Sim 1997, Franke and De Jager 2011, Schipper 2015, Fritz and Lederman 2015).

Now I agree that the classical coherence principle is false. But I think such radical weakenings of the classical picture throw out the baby with the bathwater. Let me count the ways. Firstly, this sort of move substantially undermines the force of ordinary belief-based explanations and predictions. Suppose you ask why Mary eats beans every day, and I answer that Mary thinks eating beans every day will keep her healthy. Given the regular classical view, that makes for a perfectly good explanation, assuming Mary wants to stay healthy. But on the weak version, Mary's belief by itself only yields an expectation that she will eat beans from time to time. Secondly and relatedly, doxastic dispositions are in fact about as stable as our beliefs themselves, as the classical picture says they should be. This is true for Tom's disposition to spell "dreamt" D-R-E-A-M-T: he *always* spells it that way. If he started spelling it D-R-E-A-M-E-D, you would conclude he had changed his mind about the spelling, not that his belief had failed to activate. What the classical picture got wrong was the *nature* of Tom's behavioural dispositions, not their stability. Thirdly, it is unclear, on the weakened account, why you should ever think someone lacks a belief. Say you see me calmly munch a ham sandwich. Ordinarily, you might conclude that I do not believe the sandwich is poisoned. But if I act on my beliefs only occasionally, that conclusion is not justified.

Finally, and perhaps most importantly, one of the core functions of belief attributions is to give a unifying explanation of an agent's actions across domains. For instance, suppose Bianca believes that there will be a blizzard this afternoon. This explains such varied actions as her choice of clothing, transportation, what she says to her colleagues at the water cooler, and her decision to check if her flight is on time. What is more, that same belief predictably and systematically affects Bianca's response to an infinite range of familiar and unfamiliar decision situations in which blizzards are somehow relevant. This is in large part why beliefs are so useful, and reasoning in terms of them is so powerful. Bianca tells you about the blizzard at the water cooler, and you conclude she probably took the bus today, rather than cycling to work. But if Bianca's conversation were governed by beliefs that control only her verbal behaviour, say, you would have no reason to think that. For all its faults, this is something the classical picture gets right: it predicts that Bianca's blizzard belief comes out in all these different ways.

We should be loathe to replace it with a theory that lacks that core virtue. Of necessity, we will spend much of this dissertation examining the limitations of the classical view. But as we discuss bathwater and how to dispose of it, let's not forget we also have a baby to save.

The real moral, then, is that classical coherence needs to be *replaced* with something equally systematic. The inquisitive view is that a belief is the possession of an answer to a question, and manifests itself in a general disposition to act on that answer when confronted with that question. That builds on a modified coherence principle:

> **Inquisitive Coherence:** If on some occasion, an agent acts on an answer to a question because of their beliefs, then that agent is disposed to act on that answer whenever they are faced with that question. (1.4)

Inquisitive coherence retains the stability and reliability of belief-related dispositions, and, as I shall argue in Chapter 2, it accounts for the unifying role of belief attributions as well.

1.3 Questions and Quizpositions

In this section, I want to flesh out the first conjunct of the inquisitive view of individual beliefs: belief is the possession of an answer to a specific question, rather than a general piece of information about the world. I will take it as a given that belief is a relation between agents and propositions of some kind. But instead of pieces of information, the objects of belief are *question-directed propositions*, or *quizpositions* for short. The defining characteristic of a quizposition is that it is jointly individuated by its informational component and the question it answers.

Below I will give a formal articulation of this notion using the possible worlds framework. The standard possible worlds conception of a proposition is that of a set of possible worlds, or a region of logical space $\mathscr{W}$ — call that an ***intensional proposition***. But hyperintensional conceptions of propositions can and have also been articulated within the possible worlds framework, as for instance in Groenendijk 2009 and Yablo 2014.

Our first job is to say what a *question* is. In English, the word "question" is regrettably ambiguous between a certain kind of expression, also known as an interrogative clause, and the contents of expressions of this kind. Suppose I ask Anna "When is the next U.S. presidential election?", I ask Brian "When will Americans vote for their next president?", and I ask Cécile "Quand aura lieu la prochaine élection présidentielle aux États-Unis?" Then I have asked all three a different question in the former sense, since I used different words. But in the latter sense I asked them all the same question, because these three interrogative expressions are all (approximately) synonymous. Throughout this dissertation, I use the word "question" in this latter sense only, unless explicitly indicated otherwise. In linguistic semantics (Hamblin 1958, Groenendijk and Stokhof 1984), one standard account of the content of an interrogative is as a *partition* of the space of possibilities.[2]

A ***(partition) question*** Q is a partition of logical space $\mathscr{W}$. The cells $q \in Q$ of this

[2] In principle, the use to which I am putting partition questions here is totally separate from their role in linguistics. I am certainly not relying on a particular semantics of interrogatives. Even if they had no link to natural language at all, one could still maintain that partition questions play a role in explaining behaviour. But the connection with natural language does give the inquisitive picture of belief-guided action a lot of intuitive appeal. And once we get to implementing some of these ideas into an inquisitive semantics of belief reports, the link with linguistics most definitely comes in helpful (see Yalcin 2011, §8).

> partition will be called *Q-**cells***. When two worlds w and v share a Q-cell, we write $w \sim_{\mathrm{Q}} v$. Any set of Q-cells $\mathrm{A} \subseteq \mathrm{Q}$ is an ***answer*** to Q. (1.5)

The idea here is to characterise question content in terms of the information that is required to answer the question exhaustively. Take for instance the question *How many daughters did Joel Russ have?* Its complete answers are *Russ had no daughters*, *Russ had just one daughter*, *Russ had exactly two daughters*, and so on. Together, these possibilities cover all of logical space, and because the answers are exhaustive, they are incompatible with one another. Thus they jointly form a partition D of $\mathscr{W}$. Any disjunction of complete answers to this question makes a partial answer: for instance *Russ had less than three daughters*, *Russ either had five or twelve daughters*, *Russ had an odd number of daughters*, and so on. For two possible worlds w and v, $w \sim_{\mathrm{D}} v$ just in case Russ had the same number of daughters at w and at v.

With this notion of a question in hand, the natural way to model a quizposition is as an ordered pair of a partition question and a (possibly partial) answer to that question:

> A ***question-directed proposition*** or ***quizposition*** is an ordered pair $\langle \mathrm{Q}, \mathrm{A} \rangle$, also denoted A^{Q}, the first member of which is the partition question Q that A^{Q} is said to be ***about***, and the second member of which is a Q-answer $\mathrm{A} \subseteq \mathrm{Q}$. The quizposition A^{Q} is ***true*** at a Q-cell q if and only if $\mathrm{q} \in \mathrm{A}$, and it is ***true*** at a world w if and only if $w \in \bigcup \mathrm{A}$. (1.6)

Quizpositions are discussed and axiomatised in Goodman 2018 (§7.1) and they are similar to the semantic objects found in Groenendijk 2009, Yablo 2014 and Fine 2017.

To connect this to our discussion above, let S be the partition question *How do you spell "dreamt"* and E the question *Which English words end on -MT*:

$w \sim_S v$ iff the word /drɛmt/ has the same spelling at w and v, [3]

$w \sim_E v$ iff at w and v, the same English words end on -MT

Think of the word "dreamt" (or /drɛmt/) as individuated phonetically here. The partitions S and E are very much distinct. Each cell $s \in S$ represents a possible spelling of "dreamt"; each cell $e \in E$ represents a possible list of words ending in -MT. Now let M be the set of all S-cells that represent a spelling ending in -MT, and let W be the set of E-cells in where "dreamt" is on the list. Then the quizposition M^S and W^E have the same truth-conditions, but they answer very different questions. Below, we will explain Tom's ability to reproduce the final letters of "dreamt" in terms of his belief in the quizposition M^S. At the same time, the inquisitive picture affords us the flexibility to deny that Tom believes the informationally equivalent proposition W^E, which will account for his inability to answer the sphinx.

In spoken English, focus in attitude reports appears to mark this cognitive distinction (cf. Dretske 1981, Blaauw 2013, Simons, Beaver, Tonnhauser and Roberts 2017). Intuitively, saying "Tom thinks/knows that 'dreamt' is a word ending in ***-MT*** " entails that Tom can spell the last part of the word, but not necessarily that Tom will recognise the word based on those two letters. But if you say "Tom thinks/knows that ***'dreamt'*** is a word ending in -MT," it is the opposite. (Try saying it out loud!) It has long been known that in spoken language, focus is

[3] To account for the possibility that /drɛmt/ may not have been an English word, it should perhaps be added that $w \sim_S v$ whenever the word /drɛmt/ has *no* spelling at w and v.

associated with questions or alternatives, and this corroborates the idea that a question-based analysis is the right account for this contrast.

It is worth emphasising that the particular possible worlds treatment just outlined is just one way to realise the concept of a quizposition. Alternatives are easily imagined. For instance, if your conception of a piece of information is a written sentence, then the inquisitive variant could be a pair of an interrogative and a declarative sentence, or a sentence with a marked focus. You could also integrate focus or contrast classes into an account of propositional structure to get Russellian quizpositions. The inquisitive turn can take you different places depending on your starting point.

And even if you accept definitions (1.5-6) as written, there is some amount of leeway in the interpretation of the background logical space $\mathscr{W}$. For our purposes, it will be easiest to assume throughout that $\mathscr{W}$ is the space of metaphysical possibilities. But one can add a second layer of hyperintensionality to the inquisitive picture by admitting certain well-behaved metaphysical impossibilities into $\mathscr{W}$. It may for instance be possible to make sense of worlds where unicorns exist, or worlds where Hesperus and Phosphorus are distinct, or worlds where the mathematical facts are different (see Colyvan 2000, Schwarz 2013, Rayo 2013, Bacon 2018, and also §5.6 below).[4]

---

[4] Logicians like Hintikka (1975), Levesque (1984), Priest (2005), Bjerring (2013), Smets and Solaki (2018), Berto and Jago (2019) contemplate more radically impossible worlds, at which contradictions can be true. Such worlds are not closed under classical consequence. Admitting them would make for a completely different formalism, and I think it would be difficult to formulate a cogent decision theory on this basis.

1.4 Facing Questions

The inquisitive view of belief and action connects beliefs to choices via questions. In the last section, we articulated a view of belief contents that connects beliefs to questions. In this section, I supply the missing part, and connect questions to choices. This starts with a very basic, pre-theoretical observation made in the introduction: choices confront us with questions. Setting a time for a lunch date, you need to know *What time do we both finish work?* Selecting a wine for the colloquium dinner, you wonder *Which one will make me look knowledgable?* Deciding which child to reproach, you confront the question *Who lit the curtains on fire?* And so on. The link between choices and questions pervades our ordinary thinking about decision making (see also Van Rooij 2013, Koralus 2014). And not in some hidden, subconscious way: idioms to describe hard choices in terms of asking and confronting questions exist in every language.[5]

Now that we have noticed this connection, we can ask what it consists in. Here it will be useful to look to standard developments of decision theory for inspiration, where payoff matrices are used to represent decision situations. Below is an arbitrary example, taken from Jeffrey 1990. This table represents a decision situation in which somebody is choosing whether to bring red

[5] Every language I have checked, that is. For example, "We are faced with a question" (English), "Wir standen vor der Frage" (German), "On fait face à une question" (French), "Miàn-lín yī-gè wèn-tí" (Mandarin Chinese). Similar expressions exist in Dutch, Italian, Spanish, Serbian, Turkish and Shanghainese. With thanks to my informants Vera Flocke, Simona Aimar, Andrés Soria Ruiz, Louis Rouillé, Milica Denić, a student at Bilkent University and Linmin Zhang. Another linguistic factoid in this spirit: in Dutch and Spanish, the cognates of "question" do not refer to spoken questions at all, and instead mean something like *problem* or *dilemma* ("kwestie" and "cuestión" respectively).

wine, white wine or rosé to a dinner party. Each option corresponds to a row in the table. They want to get a good match, and so the payoff depends on what turns out to be for dinner: white wine goes well with chicken or herring, but badly with beef. Red wine goes OK with chicken, better with beef and terribly with herring. And so on:

	Chicken	Beef	Herring
White	1	–1	1
Red	0	1	–1
Rosé	0.5	0	–1

TABLE 1: WINE SELECTION

Now what I want to draw attention to are the column headings of the table, representing the world states on which the outcome of this agent's actions depend. In this particular case, the three headings "Chicken," "Beef" and "Herring" are possible dinners: it could be that *there's chicken for dinner*, it could be that *there's beef for dinner*, and it could be that *there's herring for dinner*. Now these three possibilities are mutually exclusive and exhaustive (or so we pretend for purposes of reading the table). Thus they form a partition of logical space, which is to say, a *question* in the sense defined above. More specifically, this partition is the question *What will be for dinner*. Intuitively speaking, that is exactly the question this choice raises for the agent.

What a serendipitous discovery! On our way to integrating questions into decision theory, we find that they were there all along. Every major formal treatment of decision theory appeals to partitions of logical space into world states at the fundamental level (Savage 1972, Jeffrey 1990, Gibbard and Harper 1980, Lewis 1981, Joyce 1999, Ahmed 2014). These partitions just have not typically been thought of as questions, and their role in determining behaviour has been largely ignored. But really, the questions we face in confronting decisions were always hiding in the classical formalism, waiting for an inquisitive theory to seek them out and grant them their proper place in decision-making.

Formally, we will want to characterise a choice or decision problem in a way that does not build the question partition in at the outset, and then define the question raised in terms of the features of this decision problem. We will use the following definition of a decision problem:

> An ***option*** is a real-valued function $\mathbf{a}$: $\mathscr{W} \to \mathbb{R}$ from possible worlds to utility values. A ***decision problem*** Δ is a finite, nonempty set of options. (1.7)

This is quite an abstract way to represent choices. The issue what exactly it takes for a concrete choice to instantiate a decision problem Δ is complex, and in §1.6-7 below I will say more about this foundational issue. But for now, we can make do with a rough gloss: a choice instantiates Δ just in case each of the actions that the agent is choosing between is represented by some option $\mathbf{a}$ in Δ, in that the value $\mathbf{a}(w)$ accurately represents the desirability of the outcome that, at w, would have been obtained as a result of choosing the action.

An outcome matrix like table 1 above represents a particular decision problem. Each row represents a function from possible worlds to real numbers, listing its values with respect to the worlds in each column. The table is a collection of rows, representing a set of options: a decision problem as defined in (1.7) above. To enable this representation, the column partition must only group worlds together where each option takes a constant value. That is, the column partition has to be a question *addressing* the decision problem, where this is defined as follows:

> The choice Δ ***raises*** the question Q, or Q ***addresses*** Δ, just in case for every option $\mathbf{a} \in \Delta$, and every cell $q \in Q$, the outcome $\mathbf{a}(w)$ takes on a constant value for all $w \in q$. This constant value can be denoted '$\mathbf{a}(q)$'. An agent ***faces*** the question Q whenever they make a choice that raises Q. (1.8)

An intuitive gloss on (1.8) is that a question addresses a choice just in case any complete answer to the question will tell you what the outcomes of all the options would be. For instance, in the table above, knowing whether chicken or beef or herring is for dinner is all you need to know to know the payoffs of the various wine choices.

Now observe that if knowing whether chicken or beef or herring is for dinner is sufficient to know the payoffs of all the options, then knowing whether chicken or beef or Atlantic herring or Pacific herring is for dinner is *definitely* sufficient. Thus, given definition (1.8), the question *Chicken, beef, Atlantic herring or Pacific herring?*, which makes one more distinction than the tripartite question *Chicken, beef or herring?*, also counts as addressing this same decision problem. Making still more distinctions does not alter that fact, as illustrated by table 2 below. Tables 1 and 2 are just alternative representations of the same decision problem. That means that in general, multiple questions will address a given decision situation. This feature of definition (1.8) will become crucially important in Chapter 2: whenever some question Q addresses a decision problem Δ, any question that forms a more fine-grained partition than Q also addresses Δ.

	Chicken Breast	Drumsticks	Chicken Soup	Beef	Atlantic Herring	Pacific Herring
White	1	1	1	–1	1	1
Red	0	0	0	1	–1	–1
Rosé	0.5	0.5	0.5	0	–1	–1

TABLE 2: FINER DISTINCTIONS

Now consider the choice one faces when writing down the word "dreamt". Your options are all the various letter combinations you could be writing down: D-R-E-A-M-T, D-R-E-A-M-E-D, and so on. Your aim, let's say, is to spell the word correctly, and so there are only two outcomes: success and failure, 1 and 0. Which of those outcomes will result from which action depends on what the correct spelling of the word in fact is. In the actual world, the word is spelt D-R-E-A-M-T, so that writing D-R-E-A-M-T leads to success and other spellings would lead to failure. But relative to possible worlds where the correct spelling is the sensible D-R-E-A-M-E-D or the phonetic D-R-E-M-T, writing D-R-E-A-M-T would lead to failure, and some other option yields success:

	/drɛmt/ is spelt D-R-E-A-M-T	/drɛmt/ is spelt D-R-E-A-M-E-D	/drɛmt/ is spelt D-R-E-M-T	…
write D-R-E-A-M-T	1	0	0	…
write D-R-E-A-M-E-D	0	1	0	…
write D-R-E-A-M-P-T	0	0	0	…
…	…	…	…	…

TABLE 3: SPELLING "DREAMT" (Δ_{Letter})

The column headings of this table form the partition question S, *How is /drɛmt/ spelt?* So in TRIVIAL TROUBLE, we can explain Tom's spelling in the letter in terms of the fact that he has the belief M^S, that *The word /drɛmt/ ends in* ***-MT***: M^S guides Tom's choice here because it answers S, and S addresses this problem, Δ_{Letter}. (Presumably he believes M^S as part of a complete S-answer, *The word /drɛmt/ is spelt* ***D-R-E-A-M-T*** — more on that in Chapter 2.)

In his confrontation with the sphinx, Tom has a very different set of options, and the outcome of his choice depends on a different feature of the world. This time he is choosing between possible replies to the sphinx. Besides confessing ignorance, Tom could have ventured an answer, like "prompt" or "unkempt":

	Only /drɛmt/ ends in -MT	Only /ʌn'kɛmt/ ends in -MT	Only /ʌn'kɛmt/ and /prɒmt/ end in -MT	…
reply /drɛmt/	1	0	0	…
reply /prɒmt/	0	0	1	…
reply /ʌn'kɛmt/	0	1	1	…
…	…	…	…	…
confess ignorance	0	0	0	…

TABLE 4: ANSWERING THE SPHINX (Δ_{Sphinx})

In the actual world, the only winning reply is "dreamt", since that happens to be the only English word ending in -MT. But in a world with different spellings, "dreamt" would yield failure, and "prompt" success. And so on: thus Δ_{Sphinx} raises the question E, *Which English words end on -MT?* So on the inquisitive picture, the expectation is that Tom's choice in Δ_{Sphinx} is guided by his views on E. We can account for Tom's failure to produce the correct answer in terms of the fact that Tom had no beliefs about E (see also §3.4 below). Tom's view on S, though it is in some sense relevant, did not help him out here, because S fails to address Δ_{Sphinx}. Take for instance the S-cell *The word /drɛmt/ is spelt* ***D-R-E-A-M-E-D***. This entails that the reply "dreamt" would lead to failure, but it says nothing about the other replies: knowing the answer to S is nowhere near sufficient for knowing the outcomes of the options in Δ_{Sphinx}.

This contrast between Δ_{Letter} and Δ_{Sphinx} draws on objective differences in the pattern of counterfactual dependence between the available actions and their outcomes. The explanation for why Tom's belief M^S guides his choice in Δ_{Letter} but not in Δ_{Sphinx} is not that Tom happens to think of the question S only in that former situation. Absent a reason *why* he should think of S in that situation, this would not be a complete explanation of the contrast. Rather, the inquisitive theory explains the pertinence of the belief M^S to the choice Δ_{Letter} without appealing to Tom's assessment of the situation. This objectivity is essential to the predictive and explanatory power of the account.[6]

According to the inquisitive picture I want to propose, it is in some sense the *world* that interrogates the agent, not the agent who interrogates themselves. That is why the sphinx from Greek mythology makes an apt metaphor: a sphinx jumps at you from the outside, forcing you to answer her question. Of course the world only rarely dispatches flesh-and-blood messengers to interrogate us. The vast majority of questions we face are posed through other means: forks in the road, supermarket aisles and unmarked shower controls. Our replies are also mostly non-verbal: turning right, picking the Greek yoghurt, or opening the tap on the left.

[6] The question you face is independent of your interpretation of the situation, but it does depend on your mental state in at least one way. If you want to go to the baker, the crossroads confront you with the question *Which way is the baker*. But if you want to go to the butcher, the very same crossroads confront you with the question *Which way is the butcher*. The reason is that your goals and preferences go into determining how the outcomes of your choice are ranked and individuated, which affects what decision situation you are in. As discussed in §1.7 below, there may be further ways in which the decision situation one is in depends on one's mental states.

Again, I am just providing one possible articulation of that basic idea here. In particular, as we will see in §1.6, the characterisation of an option in (1.7) assumes a causal-counterfactual approach to decision theory. It also presupposes Stalnaker's logic for counterfactuals: this is needed to guarantee that for any possible world w and option $\mathbf{a}$ there is such a thing as *the* unique outcome $\mathbf{a}(w)$ that $\mathbf{a}$ would have had given the facts at w (even if $\mathbf{a}$ is not in fact chosen at w). These assumptions simplify the formalism. But they are not, I think, essential. In particular, an evidential version of inquisitive decision theory could certainly be envisaged.

1.5 Acting on Answers

Using the formal framework developed above, we can now provide completely formalised articulations of the classical and inquisitive views (1.1) and (1.2) that we started out with. The only missing puzzle piece is a precise notion of what it takes to *act* on a piece of information, or on an answer. One natural way to spell this out is using the notion of dominance:

> Suppose $\mathbf{a}$ and $\mathbf{b}$ are options, and $\mathrm{p} \subseteq \mathscr{W}$ is an intensional proposition. Then option $\mathbf{a}$ ***(strictly) p-dominates*** option $\mathbf{b}$ just in case $\mathbf{a}(w) > \mathbf{b}(w)$ for all $w \in \mathrm{p}$. An option $\mathbf{a} \in \Delta$ is ***p-dominant*** just in case $\mathbf{a}$ strictly p-dominates every other option in Δ, and is ***p-dominated*** just in case Δ contains an option $\mathbf{b}$ that strictly p-dominates $\mathbf{a}$. (1.9)

> **Individual Beliefs (Classical Formalisation)**. Belief is a relation between agents and intensional propositions. An agent believing the proposition p is always disposed to avoid p-dominated options whenever there are any. (1.10)

Note that avoiding all p-dominated options if there are any entails performing the p-dominant option if there is one. The analogous formalisation of the inquisitive view (1.2) runs as follows:

> Suppose **a** and **b** are options in a decision situation that raises Q, and A is an answer to Q. Then **a** ***(strictly)*** A***-dominates*** **b** just in case $\mathbf{a}(q) > \mathbf{b}(q)$ for all $q \in A$. (1.11)

> **Individual Beliefs (Inquisitive Formalisation)**. Belief is a relation between agents and quizpositions. An agent believing the quizposition A^Q is disposed to avoid A-dominated options in decision situations that raise Q. (1.12)

With these formal restatements of the classical and inquisitive views in place, our analysis of TRIVIAL TROUBLE can now be run entirely within the formalism. Recall Tom's letter-writing and Sphinx decision problems Δ_{Letter} and Δ_{Sphinx}, displayed on p. 28 and p. 29 respectively. Suppose we want to predict Tom's response to Δ_{Letter} on the basis of (1.10). Then we need to attribute some belief p to Tom such that writing D-R-E-A-M-T is the p-dominant course of action in Δ_{Letter}. Inspection of the table shows that this proposition p will have to entail that /drɛmt/ is spelt D-R-E-A-M-T. But if p entails this, then replying "dreamt" in Δ_{Sphinx} beats replying "I don't know" at every p-world. Given (1.10), we thus get the incorrect prediction that Tom is disposed to forego the latter action in favour of the former. Again, we find that the trouble with the classical account is that we cannot secure the right prediction about Δ_{Letter} without slipping into the wrong prediction about Δ_{Sphinx}. By contrast, if we attribute to Tom a belief in the quizposition M^S, then (1.12) correctly predicts Tom's choice in Δ_{Letter}, since Δ_{Letter} raises the question S to which that belief is an answer. But this same attribution does not commit us to a prediction about Tom's behaviour in Δ_{Sphinx}, because S does not address that decision situation.

The dominance-based formulations (1.10) and (1.12) rule out any hedging against the possibility that your beliefs could be false. This aligns with an epistemological tradition that identifies full or outright belief with absolute doxastic certainty, entailing credence 1 (Ramsey 1926, Hintikka 1962, Clarke 2013, Greco 2015a, Moss 2018, Williamson 2018). Even if it can seem a bit artificial, this simple, strong notion of full belief has substantial theoretical advantages, and that is why it will be our topic in Chapter 2. So for now, let's agree to use the word "belief" this way.

But I do not really mean to prejudge to what extent the ordinary, natural language notion of belief resembles doxastic certainty. On that matter the inquisitive picture is ultimately neutral. Once inquisitive credences are on the table, it will be obvious how to formulate inquisitive versions of various subtler, weaker accounts of belief, like the Lockean view (Foley 1992, James Hawthorne 2009; see also Leitgeb 2014, Weatherson 2016). In fact, one emergent view of weak belief even *requires* an inquisitive setting (John Hawthorne, Rothschild and Spectre 2016, Holguín 2019).

1.6 A Counterfactual Savage

The formal representation of options and decision problems introduced in §1.4 is a little bit idiosyncratic. It is a variation of the notion of an option pioneered by J.L. Savage (1973 [1954]), as a function from states of nature to outcomes. This is still the standard notion of an option used in economic theory. In this section, I compare my approach to Savage's, and explain why I am setting things up a little differently.

1.6.1 Actions and Functions

Savage took options (or "actions" as he called them) to be functions from states of the world to outcomes. This formal representation of actions is intended to capture only those aspects of an action that are relevant to deliberation, and to abstract away from everything else. It builds on something like the following picture of deliberation.

First of all, what matters to the agent about an action is its outcome, in the sense that we always act in the hope of achieving a certain sort of outcome. That outcome is distinct from the action itself. If I *turn left*, say, it is not in order to satisfy a craving for left turns, but rather because I am trying to get somewhere — to the bakery perhaps. But agents are not always in a position to make decisions based on the outcomes that their actions will have (given the way things actually are), since they do not always know what those outcomes would actually be. This is where belief enters into deliberation. For instance, the outcome of a concrete act like *turning left* is not fixed: it depends on what the world is like. In this case, it depends on where the bakery is. So the choice-worthiness of the available actions is determined by where, given my beliefs, the action is liable take me relative to where I want to go. Thus, what matters about an available action from the standpoint of deliberation is the way in which the outcome of that action depends on what the world is like. A function from possible world states to outcomes is a natural representation of this dependence.

There are three important differences between Savage's conception of an option and options as they are defined here. The first difference concerns the *outcomes* that are the output of these functions. On my definition, outcomes are real numbers, representing the amount of utility the agent might obtain as a result of their choice. In Savage's system, on the other hand, outcomes are concrete things or states that the agent obtains as a result of their action, and whose value is not specified. For instance, an outcome could be *winning a car*, *hurting a knee*, or *arriving in Barbados.* There is a reason for this difference: Savage's aim was to analyse both desire and belief simultaneously. He was therefore not content to build the agent's values and utilities into the characterisation of a decision problem, and needed a more neutral way to think about outcomes. By contrast, in this dissertation I am focussing on the analysis of belief, which puts me in a position to simplify matters by taking the utilities as given, and building them into the decision problems. (Another standard way to set aside foundational questions about utilities is to restrict attention to agents who are interested solely in monetary gain, and to decision problems where nothing but money is at stake.)

The second difference concerns the possibilities that an option takes as inputs. Savage's options are functions from coarse-grained possibilities to outcomes. But in my formalism, the options take complete possible worlds as inputs.[7] Again, there is a reason for this contrast: it is part of the present project (but not part of Savage's project) to proffer a certain analysis of what it takes for a decision to confront an agent with a particular question. So for our purposes, it will not do to take a partition of logical space into coarse-grained world states as an unanalysed, primitive constituent of the decision problem. Rather, we need to represent decision problems in a way that is neutral with regard to the way logical space is partitioned into states of nature.

---

[7] It appears Savage was sympathetic to this variant of his approach. In §5.5 of the *Foundations*, Savage suggests that from a God's-eye point of view, the true representation of an action is a function from very fine-grained possibilities to utility values — exactly as in definition (1.7).

The third difference concerns the way an action A is supposed to relate a world state S to an outcome. Savage himself is, I think, ambiguous on this point. But the standard interpretation of Savage runs as follows. The outcome of an action A given the world state S is the unique outcome O that is entailed by the conjunction *A is performed and S obtains* — *A&S* for short (see e.g. Joyce 1999, §2.2, Ahmed 2014, §1.6). So the option **a** is an accurate representation of the action A if and only if it maps every world state S to the outcome O entailed by *A&S*. Let me call this account of the relationship that actions, world states and outcomes bear to one another the *conjunctive treatment of outcomes*.[8]

The conjunctive treatment of outcomes presupposes two things about the way world states are individuated. Firstly, the world states S must be individuated finely enough, and the outcomes must be individuated coarsely enough, that the conjunction of any action A and world state S entails a particular outcome O. Secondly, the world states must be individuated coarsely enough that each action is compatible with each world state.

Now if the role of world states is to be played by a possible *world*, which are by definition individuated maximally finely, then this second condition is clearly not met. For at any possible world w, only one particular action A_w is performed, and only one particular outcome O_w is reached. So the conjunction *A is performed and w obtains* is inconsistent for all but one action A. Thus the conjunctive interpretation, straightforwardly applied, is not tenable given the definition of options as functions from possible worlds to outcomes.

---

[8] The conjunctive treatment is adopted explicitly in the decision theoretical formalisms used by most philosophers — e.g. Jeffrey 1990 [1965], Gibbard and Harper 1980, Lewis 1981, Skyrms 1982, Joyce 1999.

But, as suggested in §1.4, this is not the treatment of outcomes I intend to adopt. I want to replace the conjunctive treatment of outcomes with a *counterfactual* treatment of outcomes. According to this treatment, the outcome associated with an action A and a world state S is the outcome O such that S entails the following counterfactual: *if A had been performed, then the outcome O would have been reached* ($A \mathrel{\Box\!\!\rightarrow} O$ for short). So a Savage-style option **a** accurately represents the action A if and only if it maps each world state S to the outcome O such that S entails $A \mathrel{\Box\!\!\rightarrow} O$. In the present formalism, where the world states are possible worlds and the outcomes are utility values, this translates to the following: the option **a** accurately represents the action A if and only if it maps each world w to the utility value of the outcome O such that the counterfactual $A \mathrel{\Box\!\!\rightarrow} O$ is true at w.

This alternative treatment of outcomes relies on a substantive assumption about counterfactuals: namely that for any action A and any possible world w, there is a unique outcome O such that $A \mathrel{\Box\!\!\rightarrow} O$ is true at w. In order to secure this assumption, we need to appeal to Stalnaker's analysis of counterfactuals, and in particular to the conditional law of the excluded middle (Stalnaker 1968, 1980).[9]

[9] Let Conditional Consistency (CC) be the (relatively uncontroversial) principle that given any consistent antecedent ϕ, all consequents ψ such that '$\phi \mathrel{\Box\!\!\rightarrow} \psi$' is true must be consistent with one another. CC is validated by both Stalnaker's (1968) and Lewis's (1973) semantics for counterfactuals. Conditional Excluded Middle (CEM) is the principle that for any ϕ and ψ, either $\phi \mathrel{\Box\!\!\rightarrow} \psi$ or $\phi \mathrel{\Box\!\!\rightarrow} \neg\psi$. CEM is validated only by Stalnaker's semantics.

Now let A be any possible action and let the outcome partition be $O_1, O_2, \ldots, O_n$. Since the partition is exhaustive, CC gives us that it cannot be true that $A \mathrel{\Box\!\!\rightarrow} \neg O_i$ for every i. By CEM, it follows from this that $A \mathrel{\Box\!\!\rightarrow} O_i$ for some outcome O_i. Since the outcomes are mutually exclusive, $A \mathrel{\Box\!\!\rightarrow} O_j$ cannot be true for any other outcome O_j, again by CC. Thus CC and CEM together ensure that $A \mathrel{\Box\!\!\rightarrow} O_i$ must always hold for some unique outcome O_i.

In addition, the counterfactual treatment is aligned with causal rather than evidential decision theory. Roughly speaking, given this treatment of outcomes, the classical and inquisitive decision theories formulated in Chapter 3 end up saying that the choice-worthiness of an action A is determined by the agent's credence that A would lead to the desired outcome. By contrast, evidential decision theorists hold that what matters is instead the agent's conditional credence in the desired outcome on the assumption that the action A is performed.

Personally, I don't find either of these commitments especially bothersome. Causal decision theory is by now the default position in philosophy, and while the conditional excluded middle remains controversial, it seems to be gathering followers by the day (see for instance Higginbotham 2003, Williams 2010, Klinedinst 2011, Kratzer 2015, Goodman 2015; for the classic debate see Lewis 1973, Stalnaker 1980). Moreover, the conditional excluded middle is in particularly good standing in the present context, because the classic formulation of causal decision theory, Gibbard and Harper 1980, also relies on the principle.

Furthermore, the counterfactual treatment of outcomes has substantial benefits. Most importantly for present purposes, the treatment allows for a simple formal characterisation of options and decision problems in a way that does not build in world state partitions at the outset. But there is also a broader, independent motivation for this way of thinking about outcomes. And that is the fact that the counterfactual treatment allows us to get around a difficulty that Richard Jeffrey (1974, 1982, 1990) raised for Savage's formalism. As I will explain in the next subsection, the counterfactual treatment obviates the need for a separate stipulation that the world states be selected so as to be independent of the agent's available actions.

1.6.2 Independence

Jeffrey's puzzle may be illustrated with the dilemma of Adam, who can either choose to go to the birthday party of his sweetheart Eve, or to stay at home and study for an important exam. Adam reasons thus: "Either Eve will not love me forever, or she will. If she will not love me forever, I should study: I may still die miserable and alone, but at least I will pass that exam. If she will love me forever, then I should study too, so that I can be both happy and successful. So either way, I should study." It is widely agreed that something is wrong with the dominance reasoning Adam employs here. The reason is that Eve's love for Adam might not be unconditional, and that it may be that Eve will only love Adam if he makes the sacrifice of coming to her party in spite of his exam. And if that turns out to be the case, then partying is in fact the wiser course of action.

The trouble is that Adam represents his decision in the following way, which makes it appear that the option of studying (**s**) strictly dominates the option of partying (**p**) (here "e" stands for the possibility that *Eve loves Adam forever*):

	e	¬e
s	3	1
p	2	0

TABLE 5: STUDY OR PARTY? (ADAM'S MISREPRESENTATION)

Now from the perspective of the conjunctive treatment of outcomes, it is not immediately clear what is wrong with this representation of the choice. In particular, the conjunction that *Eve loves Adam forever and Adam studies to pass the exam* does entail that Adam gets everything he wants, while the conjunction that *Eve will not love Adam forever even though he comes to her birthday party* entails the worst possible outcome. And all that is true whether or not Eve's love in fact depends on Adam's choice. Thus the issue of whether going to the party might make Eve love him apparently becomes irrelevant to the analysis of this situation.

There is a standard fix that is made on behalf of Savage to avoid this repugnant conclusion. The fix is simply to add a constraint to the specification of a decision situation: the world states must always be in some sense independent of the action the agent will perform (this is causal or counterfactual independence for causal decision theorists, and evidential independence for evidential decision theorists; see Joyce 1999, Chs. 4-5). The trouble with the partition $\{ e, \neg e \}$ is that it does not satisfy this constraint: whether Eve will love Adam need *not* be independent of whether he goes to Eve's party or not. However, as Jeffrey argues, the imposition of this independence constraint seems objectionable. For one, it is formally inelegant and *ad hoc*. And in addition, it *prima facie* constrains the generality of the theory by restricting the theory to those decision problems where the world state happens to be independent of the options under consideration (Jeffrey 1974, 1982).

If we instead adopt the counterfactual treatment of outcomes, Jeffrey's problem no longer arises. From this perspective, the trouble with Adam's reasoning is this. In considering whether to go to the party, it is irrelevant what the world *will* be like, given how Adam actually acts. What matters is what the world *would* be like if he *were* to party and what it *would* be like if he *were* to study. It is this counterfactual outcome, the one that would have occurred, that the table entries $\mathbf{s}(\cdot)$ and $\mathbf{p}(\cdot)$ are supposed to represent. Once that is cleared up, it becomes clear that there are in fact three possibilities that need to be distinguished:

	ec	¬ec	eu	¬eu
s	1	1	3	1
p	2	2	2	0

TABLE 6: STUDY OR PARTY? (CORRECT REPRESENTATION)

Assuming Eve's love is conditional (c in the table), it is true that if Adam *were* to study he would pass the exam but lose Eve (utility 1), and also that if Adam *were* to party he would fail his exam but have Eve (utility 2). Only if Eve's love is unconditional (u in the table) do we get the payoff structure of table 5. So provided Adam assigns a reasonably high credence to c, he should judge that partying is the only rational course of action.

Thus Jeffrey's problem dissolves given the counterfactual treatment of outcomes, and we avoid the need to posit any *ad hoc* additional constraints. The consequent gains in simplicity and elegance makes the counterfactual treatment a better choice for present purposes. In fact, I think one can make the case that, for the same reason, the counterfactual understanding of outcomes makes for a more satisfactory interpretation of Savage in general.

1.7 Actions and Choices

In order for the inquisitive theory of belief and action to work as intended, we need to accept certain simple, and arguably simplified, representations of the decision situations agents find themselves in. In this section, I want to bring out and address some conceptual difficulties that arise for the theory as a consequence. These difficulties touch on tricky issues about the nature of choice and action that are beyond the scope of this dissertation. My primary concern here is to clarify what the issues are, but I will also outline some possible lines of response.

1.7.1 What are Actions?

In §1.4 above, I said that the outcome of Tom's answer to the sphinx counterfactually depended on the question *Which English words end on -MT*. But is that really quite accurate? It could be argued that other factors play a role in determining the outcomes as well. What if this happens to be a sphinx who does the opposite of what she says, rewarding Tom for the wrong answer, and punishing him for the correct answer? Or what if Tom misheard the question, and the sphinx was asking for a word ending on -NT? What if this is all a dream, and none of his actions have any real-world consequences at all? What if the sphinx is a spy working for an alien civilisation that plans to exploit trivia about English spelling in an elaborate plan to destroy the human race? And so on: the possibilities are literally endless. To make the concern vivid, note that if Tom *believed* that any of these scenarios really obtained, you would expect that to inform his decision.

The issue raises a problem for the inquisitive view of belief and action. For if we grant that all these other possibilities play a role, the outcome of Tom's choice suddenly depends on a far more complicated question: *Which English words end on -MT and will the sphinx be true to her word and what did she ask and is she an alien and is this all a dream and …* So given (1.8), it would in fact be this very complicated, highly specific question that Tom is facing in making this choice. More generally, if we take account of every possibility, then basically any concrete choice turns out to confront the agent with a highly complex question, which is moreover completely specific to the very particular situation they find themselves in.

This is problematic for two reasons. Firstly, it is implausible that agents should have antecedent beliefs about those highly complex and specific questions. This is especially true given that, as I argue in §2.7 below, our cognitive limitations prevent us from having views about very large questions (i.e. questions that make very fine-grained partitions). And if we do not have beliefs about any of the questions we actually face, then it looks as if the inquisitive view of belief and action articulated above cannot ground ordinary belief-based explanations of behaviour after all. The impact of this problem is softened by the fact that, even when we do not have the answer to a question prior to facing it, we can sometimes employ deductive reasoning and deliberation to work our way to an answer based on their existing beliefs (see §2.8 and §3.4 below). But while the possibility of deduction makes a contribution towards resolving this issue, it does not make for a complete solution. If we really take account of all the possibilities, the questions faced get so complex that it is not plausible that we can bootstrap our way to having views about them, even with the help of our deductive faculties.

The second problem is that, if we face a different question for almost every choice we make, this radically undermines the force of the Inquisitive Coherence principle articulated in §1.2. The idea that we act on a given belief when we are faced with the same question will not have any unifying force unless we can expect to face the same question on more than one occasion.

Again, the problem is less bad than it first appears: as explained in Ch. 2, beliefs on related matters turn out to be constitutively connected to one another. So inquisitive theory also unifies our actions in response to suitably related questions to some extent. But again, this may not be enough to address the full force of the objection. To address the problem at the root, I think we must resist the notion that the only "true" representation of Tom's decision situation is some hypercomplex decision table. Instead, I want to justify the representation by the relatively simple decision table Δ_{Sphinx} given above.

One way to do this is to grant that Δ_{Sphinx} is a simplified, idealised representation of Tom's choice, which fails to take account of all the possibilities. Granted that Δ_{Sphinx} is not a fully accurate representation of all the messy realities of the situation, one could insist that modelling the situation in these simplified terms is nonetheless justified. How can it be that oversimplified, inaccurate models of a situation can still yield accurate predictions and substantive explanations of real-world events? This is an interesting and complex question, but perhaps it is best left to philosophers of science. Sometimes idealisation and oversimplification are simply the price we pay for achieving at least a partial understanding of what is going on. For decision theorists, as for scientists in other fields, a measure of idealisation is just a fact of life.

A second line of response, which I am inclined to favour, is to get more specific about the nature of the actions Tom is choosing between, and which are represented by the options. Take the action of *saying "unkempt"*. There are different ways of thinking about what, exactly, it takes for this action to be performed. To see what I mean, here are some possibilities:

A) *Willing myself to produce the sound* /ʌn'kɛmt/

B) *Firing the neuron that would, under normal circumstances, lead to my vocal chords to produce the sound* /ʌn'kɛmt/

C) *Producing the sound* /ʌn'kɛmt/

D) *Articulating the English word* /ʌn'kɛmt/

E) *Replying to the sphinx by saying the English word* /ʌn'kɛmt/

F) *Replying to the sphinx, who just challenged me to give her a word ending on -MT, and who will richly reward me if I answer correctly, by saying the English word* /ʌn'kɛmt/

These descriptions of Tom's action are ordered from thin to thick. Now the solution I want to suggest is to posit that the choice represented by Δ_{Sphinx} is in fact a choice between the action (F) and other, similarly described actions. Since the performance of any one of these actions entails the basic details of the setup, we thus avoid the worry that the outcome of the action should depend on extraneous factors. (Even at worlds w where the sphinx is lying, or at which Tom is dreaming up the whole encounter, the counterfactual "if Tom *were* to perform the action (F), he would be rewarded" is still true at w if and only if at w the word /ʌn'kɛmt/ ends on -MT.)

Traditionally, decision theorists have held that the actions in decision theory are more like (A), (B) or (C), often referred to as *basic actions* (see e.g. Joyce 1999, §2.3, Danto 1965). The idea here is that the acts the agent is choosing between should always be events over which the agent has absolutely perfect control, and such that, if they were to choose one, they would be certain that that action will be realised. However, I agree with David Storrs-Fox (2019), who has argued persuasively that the notion of a basic action is a remnant of an antiquated Cartesian picture of agency, and that we have good reason to think basic actions thus conceived do not really exist (see also Lavin 2013). And if the Cartesian, internalist ideal of a basic action is unachievable anyway, we might as well go for a thick conception of action: in for a penny, in for a pound.

That being said, there is a reason most decision theorists prefer to think of options as basic actions. For if the options in a decision problem represent non-basic actions, then it may be possible for an agent to choose an option without knowing that that is the action they have chosen; and that seems to lead to unacceptable consequences. To bring this out, consider a variant of Tom's story. Suppose Tom finds himself in the situation described in TRIVIAL TROUBLE, except that he misheard the sphinx and incorrectly believes she asked him for a word ending on -NT. Consequently he replies "elephant". Tom has in fact replied "elephant" to the sphinx's question *What English words end on -MT*. But he does not know this: he thinks he replied "elephant" to the question *What English words end on -NT*.

Moreover, to account for Tom's reply, the misunderstanding obviously matters. We need to invoke his incorrect belief about what the sphinx says. And that means we have to reject the representation Δ_{Sphinx} in this case. In this situation, we need the extra columns associated with the question *What did the sphinx ask* to account for the way Tom's beliefs guided his decision-making. Furthermore, the salient difference in this variant of the story is just that Tom did not *know* that the sphinx asked for a word ending on -MT. So in order to avoid bad predictions, it looks like we will need to add a requirement to the effect that, in order for an agent to count as making a choice between a set of actions, the agent must *know* that they are making that choice. This would mean that the decision situation the agent is in depends in part on the agent's epistemic state. And that seems circular: in spelling out the action-guiding role of a belief, we end up having to appeal to some of the agent's other beliefs and knowledge.

Is this a vicious or a virtuous circularity? I confess I am not quite sure: I suspect it depends on what you want to get out of decision theory. Insofar as this is a problem, it affects decision theory generally, not just the inquisitive variant. Still, if an agent's knowledge of their situation partly determines the decision situation they find themselves in, this does warrant a further qualification of the claim from §1.4 that the *world* confronts the agent with questions. For now it seems that agents are, after all, not just the passive recipients of questions from the outside: their understanding of the situation does affect the question they count as facing.

1.7.2 What are Choices?

The problem just discussed was essentially an issue about reining in the *width* of a decision table, finding a way to deal with all the exotic possibilities one finds there. There is a somewhat similar issue to do with reining in the *height* of the decision table, and all the strange actions an agent could in principle perform. Consider again Tom's run-in with the sphinx. We represented this as a choice between possible replies to the sphinx. But it seems that Tom could have performed some other actions too; for instance, he could have decided to burst out in song, to wiggle his toes, or to fight the sphinx. One could argue, then, that all of these other possible actions should also be represented by options in the decision table.

Again, the trouble with admitting such options is that doing so will complicate the question Tom is facing. Now it looks like the question he was facing is not just *Which English words end on -MT* but rather something like *Which English words end on -MT and what songs does the sphinx like and how strong is this sphinx and…* And again, this problem is quite general: almost any decision situation can be made to seem arbitrarily complex by considering a larger number of possible actions. As before, it is implausible that anyone would have antecedent opinions about those complex questions, or even that they could form them by deduction.

Again, I think one can give two sorts of responses here, mirroring the responses floated in the last subsection. The first response is the unapologetic idealiser. This response grants that in editing down the number of options, we create an intentionally simplified representation of the situation: perhaps that is just the cost of doing business with inquisitive decision theory.

A second response seeks to exploit the wiggle room afforded by the notion of a *choice*. Maybe we should not be maximally permissive about which actions count as alternatives in a given choice: just because it was possible for the agent to perform an action does not mean that that action was an alternative in the choice they actually made. Perhaps only the actions an agent *considers* count as alternatives, say. Or there may be some other criterion for singling out a subset of the agent's possible actions as the actions the agent is choosing between. However this is spelt out exactly, such a thick conception of choices could give us a way of avoiding the problem about the abundance of actions. For it allows us to say that Tom really did make a choice in which all the alternatives were possible replies to the sphinx. The decision problem Δ_{Sphinx}, then, is a formal representation of *that* choice.

Chapter 2.
The Web of Questions

This chapter is about the way an agent's individual beliefs come together in belief states. In the classical picture, this is where the problematic rationality assumptions of the theory really become apparent. The classical view of belief states is an extreme form of holism: all of an agent's beliefs combine into a single, global view of the world. Ramsey called it "the map by which we steer" (Ramsey 1926, p. 238; Armstrong 1973; Yalcin 2018). And to fit together into this monolithic world view, our beliefs must all cohere perfectly. On the most natural development of the inquisitive view, an agent's beliefs are not as tightly knit as that. But they are not completely separate either: beliefs are still constitutively linked to each other through their thematic connections. They do not form a map, it turns out, but something much more akin to a *web* (albeit a different kind of web than the web of belief Quine had in mind).

If its view of individual beliefs is the most popular aspect of the classical picture, the classical view of belief states is probably the least popular. Even the most ardent supporters of classical decision theory tend to confess that a measure of idealisation is involved in the assumption that ordinary human beings never have inconsistent beliefs, say. So all sides can agree that there is room for improvement here, if the goal is to describe the behaviour of ordinary agents.[6] At the same time, the somewhat obvious faults of the classical view should not blind us to its strengths. Again, the inquisitive alternative aims to strike a balance, ejecting the bathwater while preserving the baby.

For instance, our beliefs are not fully closed under entailment on the inquisitive view, but we do automatically believe some of the consequences of our beliefs. For instance, you cannot believe that *"dreamt" ends in **-MT*** without also believing that *"dreamt" ends in a **T***. Here the account touches on work by Stephen Yablo (2014) and Kit Fine (2017), who developed the idea of a propositional *part*. Propositional (or quizpositional) parthood is a specific kind of entailment. And while we do not necessarily believe every consequence of our beliefs, I claim we do believe every part of what we believe (ch. 7 of Yablo 2014 makes the analogous claim about knowledge). Likewise, the other rationality assumptions of the classical picture are replaced by less demanding, more realistic inquisitive analogues.

In §2.1-2 below I will assess the classical view of belief states; §2.3-6 develop and motivate the inquisitive alternative. Both views of belief states are strongly rooted in the corresponding

---

[10] Robert Aumann (1987) once expressed the view that decision theory does not apply to agents in the actual world at all, but is solely concerned with a fictional, ideally rational species called the *homo rationalis*. Even on a less extreme outlook, some may feel that the issues raised here stretch the domain of application of the classical theory beyond its intended narrow bounds. Maybe so, but then I think decision theory *should* broaden its horizons: I agree with those who find those old bounds too constrictive, and are hopeful that behaviour that falls short of the classical ideal is still susceptible to systematic theorising (Kahneman and Tversky 1979; Thaler 2015).

views of individual beliefs. In §2.2, §2.5 and §2.11 I establish this connection. These joint results provide confirmation for the diagnosis, at the start of Chapter 1, that to fix the fundamental problems of the classical picture, we need to revisit its view of individual beliefs. In §2.6-10, I examine how, given this view, we should think about transitions from one belief state to the next, and examine one way to understand deductive inference within the inquisitive theory.

2.1 Classical Belief States and the Problem of Deduction

Recall that in formal statements of the classical theory, we take the objects of belief to be *intensional propositions* or sets of possible worlds. As is standard in decision-theoretical contexts, we will avoid complications arising from infinities by assuming that the background space $\mathscr{W}$ of possible worlds is finite. Then the classical, "map-by-which-we-steer" view of belief states can be stated as follows:

A ***classical information state*** is a set of intensional propositions $\mathbf{I}$ such that:

i) *Closure under entailment*: If p entails q and $p \in \mathbf{I}$, then $q \in \mathbf{I}$.

ii) *Closure under conjunction*: If p, $q \in \mathbf{I}$, then $p \cap q \in \mathbf{I}$.

An information state $\mathbf{I}$ is ***accurate at*** a possible world w if and only if all propositions $p \in \mathbf{I}$ are true at w; $\mathbf{I}$ is ***consistent*** if and only if it is accurate at some world. (2.1)

Belief States (Classical). An agent α's beliefs form a consistent classical information state $\mathbf{B}_\alpha$, manifesting themselves in a general disposition to forego $\bigcap\mathbf{B}_\alpha$-dominated actions. (2.2)

Defining a classical information state as a set of propositions facilitates the comparison of information states with sets of propositions that need not be closed under entailment or conjunction. But one can, without loss, represent the classical information state **I** by a single proposition $\bigcap\mathbf{I}$: **I** contains p just in case $\bigcap\mathbf{I}$ entails p, **I** is accurate at w just in case $w \in \bigcap\mathbf{I}$, and **I** is consistent just in case $\bigcap\mathbf{I} \neq \emptyset$.

So a classical agent with belief state **B** always has a strongest belief, $\bigcap\mathbf{B}$; and all their other beliefs are just consequences of this overall view of what the world is like. The possible worlds compatible with $\bigcap\mathbf{B}$ are called the agent's ***belief worlds***. The classical views (2.1-2) can be helpfully summed up in terms of this notion: agents *believe* whatever is *true* at all their belief worlds, and *do* whatever is *best* at all their belief worlds. The closure conditions (i) and (ii) immediately fall out of this slogan: if p is true at every one of your belief worlds, then so is everything that p entails. And if both p and q are true at all your belief worlds, then so is their conjunction p ∩ q.

To understand the way that, according to (2.2), the contents of a doxastic state are supposed to cohere, the analogy of a map is very helpful (see also Camp 2007, Yalcin 2018). Suppose I stick a pin into a map of New York to represent the location of Terry's taco truck. When I move the pin from Central Park to Prospect Park, the map now says that *Terry is in Prospect Park* and no longer says that *Terry is in Central Park*. But that is not the only change. Unlike before, the map now says that *Terry is in Brooklyn*, that *Terry is South of Williamsburg*, that *Terry is close to the Brooklyn Public Library*, and so on. In short, the map's contents instantly adjust to preserve overall coherence in every respect.

According to the classical picture, our beliefs should preserve their coherence in the same way. And while that may not be entirely right, there is something right about it. Suppose I think it is half past eleven. I inspect my watch and find it is a quarter past one already! Besides learning the time, I instantly lose a number of beliefs. I no longer believe that *it is half past eleven*, or that *it's morning* or that *it is not yet one o'clock*. I also acquire several other beliefs I did not have before: I now believe that *it is not half past eleven*, that *it is afternoon*, and that *it's a quarter past the hour*. It is not plausible that these are six distinct cognitive achievements, each requiring a separate act of deductive reasoning. For one, the change is apparently instantaneous. It is far more plausible that a single adjustment to my doxastic state is responsible for all these differences in the propositional content of my beliefs, just like moving the pin on the map produced a range of changes in the propositional contents of the map. The classical picture and the attendant map metaphor help us understand how that works.

More importantly for present purposes, presumptions of coherence play an important role in ordinary belief-based explanations of behaviour (see Cherniak 1986, ch. 1). Ordinarily, you would take it that my belief that it is 1.15pm is a good explanation for my affirming, when asked, that "it is afternoon". That explanation presupposes there is some connection between believing that it is 1.15pm and believing it is afternoon. In §1.1, we talked about the power of beliefs to provide a unifying explanation for our choices across domains. We discussed, for instance, how Bianca's belief that a blizzard is coming can explain what she wears, how she gets to work, what she says, and an unlimited range of other facts about her behaviour. But unless Bianca's acquisition of the blizzard belief stopped her from continuing to believe that it will be a sunny all day, it does not really explain most of those choices after all. The natural cohesion of our beliefs is a familiar fact that plays an important role in our ordinary reasoning about belief and behaviour, and the classical view of belief states does justice to this fact.

But of course, the classical view goes too far. As Frege once said, some entailments are like beams in a house, while others are like plants in their seeds (Frege 1884, §88). It is plausible that we believe the beam-in-house entailments automatically, without the need for deduction: for instance, believing Austin has a yellow car does seem to involve believing Austin has a car. But plant-in-seed entailments require effort to see. If they did not, inspecting a scrambled Rubik's cube would be sufficient for knowing how to solve it. Say you have memorised the pattern on all six faces of the cube. Then in all your belief worlds the cube is scrambled some particular way. Given what you know about the mechanics of the cube, that information entails that a certain sequence of moves *S* will unscramble the cube: in all your belief worlds, *S* unscrambles the cube. So according to (2.2), you should automatically have the ability to act on that belief by solving the cube. Likewise, logic riddles, Chinese rings, and Sudokus should all be solvable in an instant. But we all know from experience how difficult such of problems can be.

If we really knew all the consequences of our beliefs, there would be no calculation or deductive reasoning of any kind: why expend energy computing entailments that you already know? We would only acquire new beliefs when we got new information, and all the ramifications would be instantly known. Calculators would not sell, computer modelling would be unnecessary, mathematicians would be out of work. Seth Yalcin sums it up nicely: "The map picture, as developed in possible worlds style, suffers an unfortunate irony: taking the belief-desire states of agents to be constitutively rational, it makes little room for the intelligibility of deductive reasoning as a rational activity. It is hard, on this picture, to see even what deductive reasoning would be." (Yalcin 2018, p. 5)

Faced with these shortcomings of the classical picture, how are we to respond? It is tempting simply to discard the classical coherence requirements altogether. But much as with the classical coherence principle discussed in §1.2, radically weakening the theory also undermines its explanatory power. We should seek replacements for these constraints that preserve what was right about them. What might such those replacements look like? You cannot just say "beliefs are closed under obvious entailments" and be comforted. For it is a familiar fact that highly non-obvious consequences can be reached through long derivations consisting of easy steps. So before you know it, closure under obvious entailment collapses back into closure under entailment. This is a formidable challenge: for a sense of the difficulties involved, see Hacking 1967, Dummett 1975, Vogel 1990, Stalnaker 1991, 1999a, Gaifman 2004 and Parikh 2009. As we will see, one of the strengths of the inquisitive account is that it meets this challenge in a very simple and principled way.

There is another reason why getting rid of the classical rationality assumptions is easier said than done. It might appear like the rationality constraints of (2.2) are just arbitrary posits that have been bolted onto the classical view of individual beliefs. But that is not accurate. In fact, the classical view of belief states is so closely entwined with the classical view of individual beliefs that any criticism of the former must draw scrutiny to the latter. In the next section, I bring out this connection.

2.2 Derivation of the Classical View

Unlike the classical view of individual beliefs, the classical view of belief states has little *prima facie* plausibility. Its strong, explicit rationality constraints fly in the face of common sense. But those constraints are not externally imposed. They are strongly rooted in the classical view of individual beliefs (1.1), as formalised in (1.10):

> **Individual Beliefs (Classical Formalisation)**. Belief is a relation between agents and intensional propositions. An agent believing the proposition p is always disposed to avoid p-dominated options whenever there are any. (1.10)

As I will show in this section, (1.10) entails that an agent's beliefs must be consistent, and also that agents behave as if their beliefs are closed under both entailment and conjunction. So on one natural articulation at least, (2.2) merely makes explicit some rationality constraints that were already implicit in (1.1). This derivation helps set the stage for the inquisitive view of belief states, which naturally grows out of (1.2) and (1.12) in a precisely analogous way.

Let's start with closure under single-premise entailment, condition (2.1.i). Suppose Eli has some belief p, say that *it always snows heavily in January*. And let q be an arbitrary consequence of p, say that *it will snow heavily this January*. The disposition (1.10) associates with believing q is that Eli should be disposed to avoid q-dominated options whenever they arise. Now suppose some option $\mathbf{b}$ is q-dominated by an alternative $\mathbf{a}$, so that $\mathbf{a}(w) > \mathbf{b}(w)$ for all q-worlds w. Then clearly $\mathbf{a}(w) > \mathbf{b}(w)$ for all p-worlds w: since p entails q any p-world is also a q-world. And thus, given

(1.10), Eli will forego the action **b** on the basis of her belief that p. More generally, she will have every disposition associated with believing q. For instance, suppose Eli is choosing between buying a snow shovel or a set of deck chairs. If *snow this January* is by itself enough to establish the snow shovel is the superior purchase, then *snow every January* is more than enough. And so given (1.10), she will pick the shovel over the deck chairs just on the basis of the stronger belief. More generally, (1.10) predicts that Eli behaves in every respect as if she believes that *there will be snow this January*: she will assert it when questioned, prepare for the snow, etcetera. At that point we might as well say she *does* believe it. Generalising one step further, (1.10) entails that Eli behaves as if she believes anything entailed by one of her beliefs. In this way, (1.10) naturally leads into closure under entailment.

The derivation of condition (2.1.ii), closure under conjunction, is a bit more involved, and requires us to understand (1.10) as covering *composite* as well as simple choices (see §4.3 below for details). But the basic thought is the same: given (1.10), believing p and believing q entails that you act as if you believed their conjunction. From the point of view of describing your behaviour, then, we might as well say that you *do* believe the conjunction. It is not at all obvious that believing p and believing q should entail a disposition to avoid pq-dominated options: after all, an option can be pq-dominated without being either p-dominated or q-dominated. But the entailment does in fact hold. For we can show that any agent who is *not* disposed to avoid pq-dominated options can find themselves in a situation where they will choose either a p-dominated or a q-dominated option no matter what they do. So in order to be able to avoid both p-dominated and q-dominated options in general, you have to avoid pq-dominated options as well.[11]

Finally, there is (2.2)'s requirement that the agent's beliefs be consistent. Well, suppose an agent has inconsistent beliefs $p_1, p_2, \ldots, p_n$. As we saw, (1.10) entails that this agent behaves as if they believed their conjunction $p_1 \cap p_2 \cap \ldots \cap p_n$, which in this case is a contradiction. Thus, given (1.10), any agent with inconsistent beliefs would be disposed to avoid ⊥-dominated options. But what does that even mean? Suppose our inconsistent agent arrives at a fork in the road and can go either left or right. Since ⊥ entails that going left dominates going right, (1.10) predicts they should go left and not right. But ⊥ also entails that going right dominates going left, so (1.10) also predicts the agent goes right and not left. But of course the agent cannot do both. Thus (1.10) entails no agent can have inconsistent beliefs, since any such agent would have to have impossible dispositions.

Jointly, these three arguments show that (1.10) entails that any believer behaves *as if* their beliefs

---

[11] Suppose an agent is *not* disposed to avoid pq-dominated options. That is, suppose they will at least sometimes choose an option **b**, even though there is an alternative **a** such that $\mathbf{b}(w) < \mathbf{a}(w)$ for all $w \in pq$. And suppose that immediately after having chosen **b** over **a**, this agent is offered a bet; if they refuse, they are guaranteed to get 0 utility (call that option **o**). But if they **t**ake the bet, their payoffs are as follows:

$$\mathbf{t}(w) = \begin{cases} \frac{1}{2} \cdot (\mathbf{a}(w) - \mathbf{b}(w)) & \text{for all } w \in pq \\ 1 & \text{for all } w \in p\neg q \\ -1 - |\mathbf{a}(w) - \mathbf{b}(w)| & \text{for all } w \in \neg p \end{cases}$$

If our agent avoids p-dominated options, they will take the bet, because **t** strictly p-dominates **o**. But note that the composite choice of **b** and **t** is strictly q-dominated by the composite choice of **a** and **o**:

$$\mathbf{b}(w) + \mathbf{t}(w) = \begin{cases} \frac{1}{2} \cdot (\mathbf{a}(w) + \mathbf{b}(w)) < \mathbf{a}(w) & \text{for all } w \in pq \\ \min\{\, \mathbf{a}(w), \mathbf{b}(w) \,\} - 1 < \mathbf{a}(w) & \text{for all } w \in \neg pq \end{cases}$$

So if they take the bet, the agent has failed to avoid a q-dominated option. But leaving the bet is a p-dominated option. Either way, having chosen **b** over **a** the agent cannot avoid both p-dominated and q-dominated options.

formed a consistent, classical information state. That still leaves a little bit of light between the classical view of individual beliefs (1.10) and the classical view of belief states (2.2), which says that an agent's beliefs actually *do* form such a state. One way to bridge this gap would be with something like the following principle:

> **Quacks Like a Duck Principle.** If an agent α has the behavioural dispositions that are associated with a belief with a certain content, and moreover α has those dispositions in virtue of their beliefs, then α actually does have a belief with that content. (2.3)

That is, if they look like they believe p, swim like they believe p and quack like they believe p, and moreover do all those things in virtue of what they believe, then the agent probably believes p.

By rejecting (2.3)'s moderate behaviourism, you create a little bit of space between (1.10) and (2.2). However, while there may be reasons for rejecting (2.3), doing so will not be of any help in addressing the problems that concern us in this dissertation. The sorts of behavioural patterns we want to account for here show that the agents in question do not in fact have classical belief states, which conflicts with (2.2). But they also show, even more directly, that contrary to (1.10), agents do not act as if they had classical belief states either. No acceptable classical belief state accounts for their choices in accordance with (2.2). Ordinary human beings, I contend, do not look classical, swim classical or quack classical. The problem is to provide some other belief-based explanation of their choices. The derivation from this section establishes that any such explanation must reject (1.10), which entails that agents do behave classically.

2.3 Quizpositional Mereology

Just like intensional propositions, quizpositions stand in entailment relations: A^Q entails B^R just in case B^R is true at every possible world where A^Q is true (that is, just in case A^Q *necessitates* B^R or just in case $\bigcup A \subseteq \bigcup B$). But in addition, quizpositions also stand in a stronger consequence relation known as *parthood* — this is a notion developed by Yablo (2014) and Fine (2017). Quizpositional parthood is essentially a particularly blatant or obvious sort of entailment. To invoke Frege's metaphor once more, there are beam-in-a-house and plant-in-a-seed entailments. The parts of a quizposition are its beam-in-a-house entailments. The paradigm example of a part is a conjunct, and the paradigm example of a non-part is a disjunction. For instance, *It's Friday* and *I'm very tired* are both part of *It's Friday and I'm very tired*. But *Either it's Friday or I'm very tired* is not part of *It's Friday*. To introduce the concept of quizposition parthood, I will begin by saying something about question conjunction and question parthood.

In natural language, interrogatives can be conjoined just as declarative sentences can. Consider for instance the conjunctive question *How many daughters did Russ have and how many sons?* A complete answer to that question requires a complete answer to each conjunct. That is, any complete answer is the conjunction of some complete answer to D, *How many daughters did Russ have?*, and a complete answer to S, *How many sons did Russ have?* Thus the cells of the conjunction of D and S, written DS, are the intersections of D-cells with S-cells. In general,

The ***conjunction*** of two questions Q and R, written QR, is the question

$$QR := \{ (q \cap r) : q \in Q \text{ and } r \in R \} \setminus \{\varnothing\}$$

> Equivalently, QR is the partition such that $w \sim_{QR} v$ if and only if $w \sim_Q v$ and $w \sim_R v$. (2.4)

It is worth noting that not every incomplete DS-answer can be expressed as a conjunction of a D-answer and an S-answer: for instance the DS-answer *Russ has more daughters than sons*.

To obtain a notion of parthood for questions, we abstract away from the relation that question conjuncts bear to their conjunctions. Note that D is a *coarsening* of DS: each D-cell is a union of smaller DS-cells. Likewise, S is a coarsening of DS. This suggests the following definition of a question part:

> One question Q ***contains*** (or is ***at least as big as***) another question R if and only if every R-cell is a union of Q-cells. We say R is ***part of*** Q if and only if Q contains R. Equivalently, R is part of Q just in case $w \sim_R v$ whenever $w \sim_Q v$.[12] (2.5)

Note that Q contains R if and only if QR = Q. More intuitively, one question is part of another if it needs to be resolved to get a complete answer to the bigger question. For example, *What month is it?* is part of *What is the date?* and *What street does she live on?* is part of *What is her address?*

Figure 1 below illustrates the definitions visually. Each square represents a different partition of logical space: a different question. The questions in the higher tiers partition the space of possibilities more finely; they make more distinctions between possible worlds. Thus they

[12] Somewhat confusingly for our purposes, linguists refer to this relation between questions as *entailment*.

contain the smaller questions displayed below them. The top question is the conjunction of all the questions below it, since it makes every distinction those questions make and nothing besides. Likewise, each question in the middle tier is the conjunction of the two polar questions immediately below it on the bottom tier. The conjunction of some questions is always the smallest question that makes all the distinctions of its conjuncts.

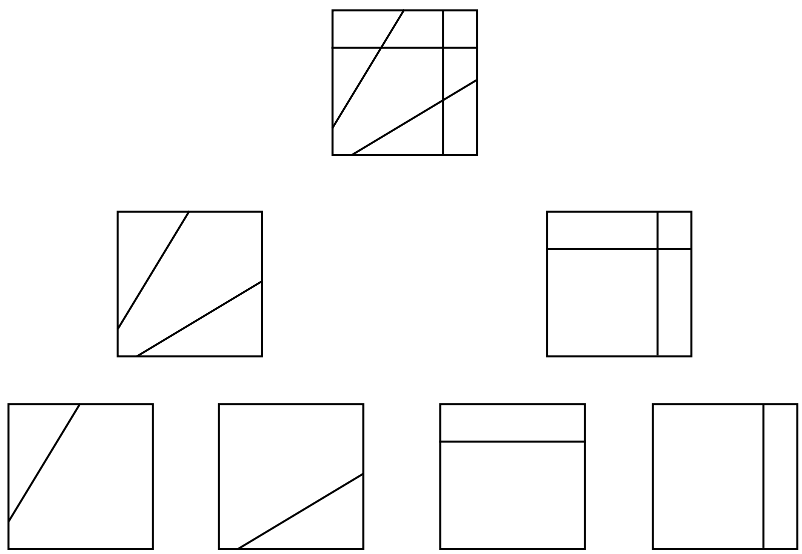


FIGURE 1: QUESTION PARTS AND QUESTION CONJUNCTION

Any conjunction of parts of Q is itself a part of Q. It follows that the common parts of any two given questions must be closed under conjunction, so that there must always be a greatest common part:

> The ***overlap*** of two questions Q and R is the biggest question that is part of Q and also of R. Two questions ***overlap*** if and only if their overlap is not equal to $\{\top\}$. (2.6)

For instance, the question *What are the capitals of Europe?* overlaps *What are the capitals of Asia?*, and their overlap is the question *What are the capitals of Turkey and Russia?* (assuming for the sake of the example that it is not contingent what the countries of Asia and Europe are). Any answer to the latter question is a partial answer to both of the bigger questions. The more (partial) answers two questions have in common, the more they overlap, and two questions do not overlap at all just in case they do not have any contingent partial answers in common.

Our definition of question conjunction naturally suggests a definition for quizposition conjunction:

> The ***conjunction*** of a Q-answer A and an R-answer B, written AB, is the QR-answer $\{(a \cap b) : a \in A \text{ and } b \in B\}\backslash\{\varnothing\}$. The conjunction of the quizpositions A^Q and B^R, written AB^{QR} or $A^Q \wedge B^R$, is the quizposition $\langle QR, AB\rangle$. (2.7)

A quizpositional conjunction makes just enough distinctions between possible worlds to make every distinctions made by its conjuncts, and rules out just enough possibilities to rule out every possibility ruled out by its conjuncts. The notion of quizposition parthood is, again, a generalisation of the relationship quizpositional conjuncts bear to their conjunction. One quizposition is part of another if it makes fewer distinctions *and* rules out fewer possibilities:

> A quizposition A^Q ***contains*** a quizposition B^R if and only if Q contains R and A^Q entails B^R (that is, $\bigcup A \subseteq \bigcup B$); alternatively, we can say B^R is ***part of*** A^Q. (2.8)

As in the case of questions, one quizposition contains another just in case the conjunction is equal to the whole. That is to say, A^Q contains B^R if and only if $AB^{QR} = A^Q$.

For example, consider the quizposition F^T, *It is* ***the 15th of January***, a complete answer to the question T, *What is today's date?* In order to answer T, you have to answer its component questions M, *What month is it* and D, *What day of the month is it*. And if your answer to T is to be F, you have to answer those component questions a particular way. Thus the quizpositions J^M, *It is **January*** and the quizposition X^D, *It is the **15th** day of the month*, are both part of F^T. Here are a few more examples to get the gist of it: *Romeo's number starts in a **two*** is part of *Romeo's number is* ***212-529 6300***; *Arden lives on **Broad Street*** is part of *Arden lives on **15 Broad Street***; and *Leopards run **very fast*** is part of *Leopards and tigers run **very fast***. In each case, the part is an answer to a small question that is incorporated in the whole, which answers a bigger question.

In the example above, J^M is in fact a somewhat special part of F^T, in that J^M is the strongest answer to M that F^T entails. More partial answers to M, such as *It is **either January or March*** or *It is **not yet August***, also count as parts of F^T. But they are not *maximal* M-parts:

> Let A^Q be any quizpositions, and R any part of Q. Then the ***maximal R-part*** of A^Q, written A/R, is the strongest part of A^Q about R. (2.9)

For instance, $J^M = F/M$ and $X^D = F/D$.

In figure 2 below, each square represents a quizposition. The inquisitive component is represented as a partition as before, and the informational component is represented as a colouring: cells where the quizposition is true are coloured light green and those where it is false a dark red. The quizpositions in higher tiers contain those displayed directly below them: their inquisitive components are more fine-grained, and they are true in fewer worlds. In fact, each part displayed in figure 2 is the maximal part about its question: that is to say, if we name the top quizposition A^Q, the others are all of the form A/R for different parts R of Q. The quizposition at the top is the conjunction of the quizpositions below it. Likewise the quizposition on its lower left is the conjunction of the two quizpositions immediately below that. But the quizposition on the right in the middle tier is strictly stronger than the conjunction of the quizpositions below it: a conjunction of two necessary truths would itself be a necessary truth.

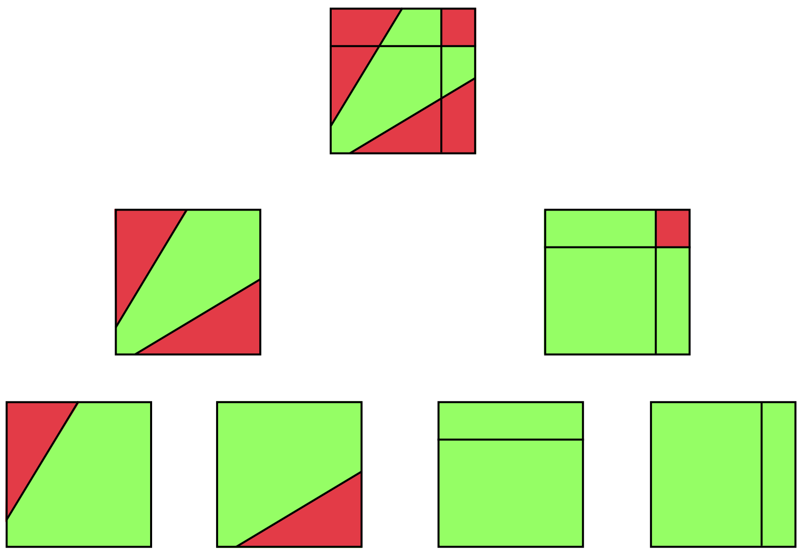


FIGURE 2: QUIZPOSITIONAL PARTS

I should note explicitly that quizposition and question parthood are both transitive relations: parts of parts of a whole are themselves also parts of the whole. For if Q makes more distinctions than R, which in turn makes more distinctions than S, then Q makes more distinctions than S. And if A^Q rules out more possibilities than B^R, which in turn rules out more possibilities than C^S, then A^Q rules out more possibilities than C^S. This is important, because it addresses the worry from §2.1 that in closing beliefs under obvious consequence we might end up ruling in non-obvious consequence too. Closing under parthood, no such collapse can occur.

Take an arbitrary set of quizpositions **S** and let $\mathbf{P_S}$ be the set of parts of quizpositions in **S**. It follows from transitivity that $\mathbf{P_S}$ already contains its own parts. There is no risk of finding any hidden, non-obvious parts amongst the parts of the parts of **S**.

Given the proposal to think of beliefs as answers to questions, closure under parthood is a very intuitive constraint. We noted in §2.1 that beliefs are not closed under entailment. Yet someone who can tell you the date can always tell you the month, and anyone who can tell you my phone number can also tell you its first digit. Clearly this is no coincidence. The quizposition mereology just developed tracks the intuitive explanation for those regularities: telling you the month is *part* of telling you the date, and telling you the first digit is *part* of telling you the phone number; that is why in each case the latter ability entails the former. In general, beliefs are answers to questions according to the inquisitive picture, and it is part of answering any big question to answer its component questions.

Conversely, your views about big questions incorporate your beliefs about their parts. For instance, if you firmly believe *It is the **20th*** in answer to the question *What day of the month is it*, then you cannot at the same time be unsure whether *It is the **20th or the 21st April*** in answer to the question *What date is it*. We can capture that idea with a partial conjunctive closure constraint: if you answer A to Q, and you answer B to some part R of Q, then B must be part of your view of Q as well, and thus you believe the conjunction AB^Q.

Recall that a classical information state is a set of intensional propositions closed under entailment and conjunction. An inquisitive information state is a set of quizpositions subject to weaker closure conditions:

> An ***inquisitive information state*** is a set of is a set of quizpositions $\mathbf{I}$ subject to the following closure conditions:
>
> i) *Closure under parthood*: If A^Q contains B^R and $A^Q \in \mathbf{I}$, then $B^R \in \mathbf{I}$.
>
> ii) *Partial closure under conjunction*: If Q contains R, and A^Q, $B^R \in \mathbf{I}$ then $AB^Q \in \mathbf{I}$.
>
> An inquisitive information state $\mathbf{I}$ is ***consistent*** if and only if there is a possible world at which all quizpositions in $\mathbf{I}$ are true. $\mathbf{I}$ is ***coherent*** just in case it contains no quizposition of the form $\bot^Q$. (2.10)

Partial closure under conjunction guarantees that whenever $\mathbf{I}$ contains any quizpositions about Q at all, $\mathbf{I}$ contains some strongest quizposition V^Q about Q. Thus we can define:

> The ***domain*** of $\mathbf{I}$, denoted $\mathscr{D}_\mathbf{I}$, is the set of all questions about which $\mathbf{I}$ contains at least one quizposition. For any $Q \in \mathscr{D}_\mathbf{I}$, $\mathbf{I}$'s ***view*** on Q, denoted $\mathbf{I}(Q)$, is the strongest Q-answer V such that $V^Q \in \mathbf{I}$. (2.11)

As the notation suggests, instead of modelling inquisitive information states as sets of quizpositions, it is also possible to think of them as a special kind of function from questions to views.[13] Thus inquisitive information states are a more constrained versions of the inquisitive belief models Yalcin (2011, 2018) proposed — more on the similarities and differences with Yalcin's view in §3.5 below.

---

[13] It is a function $\mathbf{I}$ from questions to views about those questions subject to two conditions: (a) $\mathbf{I}$'s domain $\mathscr{D}_\mathbf{I}$ is closed under question parthood, and (b) when Q, R $\in \mathscr{D}_\mathbf{I}$ have a common part S, $\mathbf{I}(Q)/S = \mathbf{I}(R)/S$ — see (2.14) below.

2.4 Inquisitive Belief States and the Possibility of Deduction

We now have everything in place to state the inquisitive view of belief states:

> **Belief States (Inquisitive)**. An agent's beliefs form a coherent inquisitive information state $\mathbf{B}_\alpha$, manifesting themselves in a disposition to forego $\mathbf{B}_\alpha(Q)$-dominated options when confronted with any question $Q \in \mathscr{D}_{\mathbf{B}_\alpha}$. (2.12)

In this section, I will examine some simple choice situations where this view fares better than the classical view of belief states (2.2) — in particular, we will be interested in the way the inquisitive picture leaves room for deductive failure and inconsistencies. In §2.5 and §2.11, I will discuss the internal motivation for the inquisitive view of belief states (2.12), showing how to derive it from the view of individual beliefs (1.12), in much the same way that, in §2.2, we derived the classical view of belief states (2.2) was from the view of belief states (1.10).

On the inquisitive view, our beliefs form a coherent inquisitive information state. To begin motivating that thesis, consider a case that causes trouble for the classical view, taken from the behavioural economics literature (Kahneman and Frederick 2002): [14]

> MITTEN STATE MURDERS I: You ask Mandy, a smart Arizona criminology major, how many murders took place in Michigan last year. After some hesitation, she guesses

[14] In Kahneman and Frederick's experiment, a large group students from the University of Arizona was asked to estimate the yearly murder rate in Detroit, and another large group was asked to estimate the yearly murder rate in Michigan. The median response for "Detroit" was 200, for "Michigan" it was 100. (The actual rates are higher.)

> "around 150". Then you ask "What about *Detroit*?", to which she confidently replies: "How silly, I didn't think of Detroit! Detroit alone had well over two hundred murders last year, so Michigan's number must be even higher."

First we might try to analyse this example from a classical viewpoint. At the beginning of this story, what is the world like according to Mandy? To explain her confident reply to the follow-up question, it seems we need to say that Mandy knew all along that there are over two hundred murders a year in Detroit, and also that she knew Detroit was in Michigan. Clearly she did not learn either fact over the course of the conversation: she had this information already. But on the classical picture, we are then forced to say Mandy believed from the start that there are over two hundred murders a year in Michigan. For if all of Mandy's belief worlds are worlds where Detroit, with its 200+ murders, is in Michigan, then there are also 200+ Michigan murders in her belief worlds. But given *that*, (2.2) would predict, wrongly, that Mandy should give an answer to that effect. So no matter what classical belief state we attribute to Mandy, we are unable to explain both her replies.

For an inquisitive account of Mandy's first reply, we do not ask for her world view in general. Instead we look for a view about M, *the number of murders in Michigan last year*:

$w \sim_{\mathsf{M}} v$ iff the number of murders in Michigan last year at w is the same as at v

We can account for Mandy's tentative reply if we say she has no strong antecedent beliefs about M. Maybe she lacks beliefs about M altogether, or maybe she has some weak view about M: she may believe, say, that there were fewer than a thousand murders. But evidently, she does *not* yet believe A^{M}, that *Michigan had* ***over two hundred murders*** *last year*. To account for Mandy's second reply, we need to appeal to her view on a different matter D, *the number of murders in*

Detroit last year:

$$w \sim_{\mathrm{D}} v \quad \text{iff} \quad \text{the number of murders in Detroit last year are the same at } w \text{ and } v$$

We can explain Mandy's assertion about Detroit by saying she believes $\mathrm{B^D}$, that *there were **over two hundred** murders in Detroit last year*. If, for simplicity, we ignore the possibility that Detroit is in a different state, $\mathrm{B^D}$ entails $\mathrm{A^M}$. But Mary can nevertheless believe $\mathrm{B^D}$ without believing $\mathrm{A^M}$ because these quizpositions are about disjoint questions: D and M do not draw any of the same distinctions between worlds, and so they do not overlap. Thus the inquisitive account has a straightforward and intuitively satisfying explanation for Mandy's behaviour.

At the start of MITTEN STATE MURDERS I, Mandy believes $\mathrm{B^D}$ but not $\mathrm{A^M}$. By the end, Mandy has put two and two together. She has made a deductive inference, and as a consequence she now believes both $\mathrm{B^D}$ and $\mathrm{A^M}$. That is why, if someone were to ask her about the Michigan murder figures again, she will not make the same mistake a second time. We can make sense of this cognitive change because, as noted in §1.1, on the inquisitive account, one can acquire a new belief without getting any new information. (In §2.6 and §2.8 below, I will provide a more detailed account of how to understand Mandy's deductive step in this example.)

Failures of deductive closure can easily lead to inconsistencies. This can be illustrated with an alternative ending to Mandy's story:

> MITTEN STATE MURDERS II: After Mandy guesses that there were around 150 murders in Michigan, you reply: "That is right! In fact, with only 120 murders, Michigan has one of the lowest murder rates of any state." Mandy unreflectively takes your word

for it, and repeats the statistic to a fellow student later that day.

At the end of this version of the story, Mandy has acquired the belief C^M, that *Michigan had **120** murders last year*, and is acting on it. If she had made the link with Detroit, she may not have accepted C^M so easily. It may be that when her fellow student mentions Detroit, Mandy will realise you had given her false information, and reject C^M. But to account for that, we need to say that Mandy retained the belief about Detroit she had in the beginning, B^D. And thus Mandy now has inconsistent beliefs: there is no possible world at which B^D and C^M are both true.

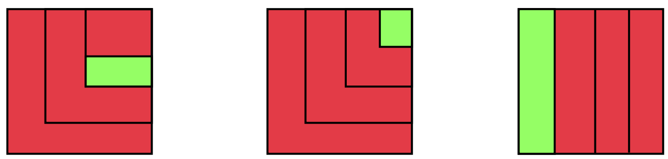


FIGURE 3: PAIRWISE INCONSISTENT YET COMPATIBLE BELIEFS

On the classical view, inconsistent belief states are ruled out, but the inquisitive view permits them. Figure 3 illustrates the situation visually: (2.12) allows for arbitrary inconsistencies between beliefs about non-overlapping questions like M and D; views about overlapping questions can be inconsistent too, provided they agree on the overlapping part (more on that in §2.11 below).

Note that Mandy's doxastic state at the end of MITTEN STATE MURDERS II violates all three classical coherence requirements: it is inconsistent, and it is neither closed under entailment nor under conjunction. But while Mandy's beliefs are inconsistent, they are not *incoherent*: Mandy does not believe any outright contradictions or necessary falsehoods. This is possible because her beliefs are not fully closed under conjunction. Mandy believes *Michigan had **120** murders last* *year*. She also believes that *Detroit had **over 200** murders last year*. But at no point does Mandy accept their conjunction, *Michigan had **120** murders last year even though Detroit had **over 200***.

2.5 Inquisitive Holism

Failures of deductive closure motivate a weakening of the classical coherence constraints. But why weaken them in this specific way? The aim of this section is to build a better understanding of what the inquisitive view says, and why it says what it does. By the end, we should begin to see why the imagery of an interconnected *web* of beliefs is an appropriate description. But I want to start by explaining some of the internal reasons for adopting the specific coherence constraints of (2.12), given the inquisitive picture developed in Chapter 1. These arguments are informal versions of those in §2.11, where I give a derivation of (2.12) from the inquisitive view of individual beliefs (1.12), along the same lines as the classical derivation in §2.2.

To begin with, let's focus on an agent's beliefs about a particular question Q. All Q-beliefs govern the same class of choices: namely all and only those decision situations that Q addresses. Because of this your Q-beliefs have to cohere with other Q-beliefs just as in the classical case, even if they might cohere less tightly with your beliefs about other questions. For the same reason that a classical agent's beliefs must cohere into a view about the world, an inquisitive agent's answers to Q will cohere into a view V^Q about Q, and their beliefs about Q are all and only those quizpositions A^Q entailed by this view V^Q.

So instead of a single monolithic world view that incorporates all of their beliefs, an inquisitive agent has different views encapsulating their answers to different questions. They will not have a view on every question: on some questions they may have no beliefs at all. But whenever an agent has answers to Q, there is always a strongest Q-answer V amongst them, and this is the agent's view on Q. Just like in the classical case, the view V^Q must be consistent. After all, every option is dominant given $\bot^Q$, but an agent can only choose one option when faced with Q. That gives us (2.12)'s requirement that belief states must be coherent.

Things get more interesting when thinking about the inquisitive closure conditions, which require us to investigate the relationship is between an agent's views on different questions. The crucial observation here was made in §1.4: if a question R is big enough to address a certain decision problem Δ, then it follows from definition (1.8) that any question Q that makes more distinctions than R also addresses Δ. In other words:

> If the question R is part of the question Q, then Q addresses every decision problem that R addresses. (2.13)

From (2.13), we can infer that if Q contains R, all the choices that are guided by the agent's view about R are also guided by their view about Q if they have one. And that fact imposes some constraints on the relationship between the agent's view on Q and their view on R.

For concreteness, let Q be the question *What are the two biggest cities in Brazil, in order?* And suppose our agent, Nico, has a view on Q, namely that ***Rio and São Paolo*** *are the two biggest cities in Brazil*, (Nico takes no stance on which of the two is bigger). How does this view constrain his views about a smaller questions? In particular, what does it tell us about Nico's view on R, *what is the biggest city in Brazil*? In the diagram below, A^Q represents Nico's view about Q. So roughly speaking, the green cells are the possibilities Nico considers *live* for the purpose of practical deliberation. The red cells are the possibilities he *rules out*. There are two Q-cells Nico considers live, namely *Rio is the biggest and São Paolo the second-biggest* and *São Paolo is the biggest and Rio the second-biggest*. The quizpositions B^R, C^R, D^R and E^R are possible views on R, and we will investigate which of these views could in principle be Nico's.

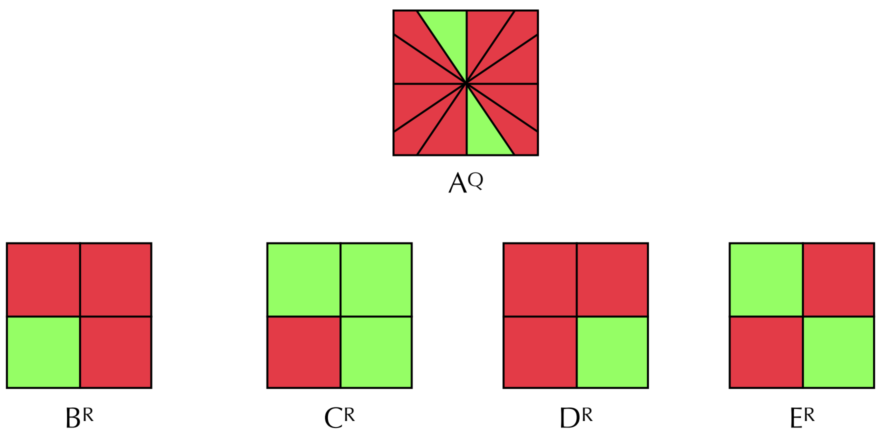


FIGURE 4: WHICH VIEWS OF THE PART FIT THE VIEW OF THE WHOLE?

First of all, Nico's view on R must be consistent with A^Q. It cannot be, say, B^R, that *the biggest city in Brazil is* ***Salvador***. For suppose you ask Nico whether or not Salvador is the biggest city Brazil. Assume Nico wants to answer truthfully, and is choosing between "Yes" and "No". Both Q and R address this choice. Given that he has the view A^Q, Nico rules out the possibility that Salvador is the biggest, and will therefore answer "No". But if he believed B^R, Nico would rule out the possibility that Salvador is *not* the biggest, and would answer "Yes". He cannot do both, so Nico cannot possibly have both views: they are associated with incompatible dispositions.

More generally, it is not possible for Nico's views about R to be incompatible with his view about Q.

Next up is C^R, *the biggest city in Brazil is **not Salvador***. Could that be Nico's view on R? C^R is consistent with A^Q, but it leaves open more R-cells than A^Q does. In particular, C^R leaves open the possibility that Brasília is the biggest city in Brazil. Whenever Nico is faced with the question R, he also faces Q, and so just in virtue of his view A^Q Nico will rule the possibility that Brasília is the biggest. But if C^R were Nico's view, then he ought to consider this a live possibility at least some of the time. He does not, so Nico's view on R must in fact be stronger than C^R. More generally, Nico must rule out every R-possibility incompatible with A. Consequently, Nico believes every part of A^Q about R.

In fact, this reasoning did not rely on special assumptions about Nico or about R, so we can generalise and conclude agents believe every part of their views. That gives us (2.10)'s closure under parthood. One can restate this condition as follows: if R is part of Q, then your view of R rules out every R-cell that your view on Q rules out. And thus everything that your view on Q says about R is also part of your view on R. (2.10)'s partial conjunctive closure condition is basically the inverse of this: provided you have beliefs about both Q and R, your view on Q must rule out every Q-cell that is inconsistent with your view on R. In other words, your beliefs about R are part of your view on the bigger question Q.

To motivate this condition, we need to show that Nico's view on R cannot rule out *more* possibilities than A^Q does. For example, Nico cannot believe D^R, that *the biggest city in Brazil is*

São Paolo. The exact reasons for this incompatibility are rather subtle and best left until §2.11. For now, let me just give an intuitive example to get at the basic idea. Suppose Nico is participating in a quiz show and is asked *What is the largest city in Brazil?* If he gets the right answer, he will win $1000. For $200 he can buy an extra guess. So he can either hedge and give two answers, or he can decide not to hedge and give one answer. Both Q and R address his choice. Nico's view A^Q leaves open two live possibilities: "Rio" might be the correct answer or it might be "São Paolo". Based on that, you would expect that Nico will play it safe: buy the extra guess, answer "Rio or São Paolo", and walk away with a sure $800. But given D^R, going all in for "São Paolo" is the strictly dominant option. So if Nico believed D^R, he would not hedge, while his view A^Q led us to expect he would hedge. He cannot do both, indicating there is at least a tension involved in combining the view A^Q with D^R.

We have now motivated all three of (2.10)'s coherence constraints. If we combine the last two, the closure conditions, we get that Nico's view on R rules out neither *fewer* nor *more* R-cells than his view on Q does. This leaves only one possibility, namely the view that rules out all and only those R-cells that A^Q rules out. This is A/R, the maximal R-part of A^Q (df. 2.9). In this case it is the view E^R, *the biggest city in Brazil is* ***either Rio or São Paolo***. So in general, an inquisitive agent's view about a big question fully determines their view about every part of that question in this way: if R is part of Q and the agent's view on Q is V^Q, then their view on R must be V/R.

It immediately follows from this conclusion that when an agent has views on overlapping questions, those views will agree on the shared part. That is to say:

If $\mathbf{I}$ is an inquisitive information state, and two questions Q, R $\in \mathcal{D}_{\mathbf{I}}$ have a common part S, then $\mathbf{I}(Q)/S = \mathbf{I}(R)/S = \mathbf{I}(S)^S$ (2.14)

Say you have a view on *What the capitals of Europe are* and also a view on *What the capitals of Asia are*. Then it follows from (2.12) that those two views must agree on the capitals of Turkey and Russia. You cannot have a firm view about the capital of Turkey with respect to the first view, and be neutral about it with respect to the latter. So certain changes in view about *What the capitals of Europe are* directly affect your view on *What the capitals of Asia are*, even though neither view is part of the other. More specifically, this will happen whenever the agent changes their mind about a question that is in the overlap of those two questions. That means you can think of your view on the capital of Turkey as quite literally being a shared part between these two views; when you change the shared part, you thereby change both the wholes it belongs to.

(2.14) does not only link views that bear a direct mereological relation to one another. Two views can overlap with a third view without overlapping each other. And even when that does not happen, they might still be connected to one another through one or more daisy-chains of intermediate views, where each link in the chain overlaps with its neighbours. As I discuss in the next section, changes in view at one end can sometimes be felt throughout the chain.

An ordinary human being has a vast number of views on all sorts of interrelated matters. So their views can be expected to form a highly complex mereological structure. For lack of a better metaphor, this is naturally thought of as a large and messy web: hence "the web of questions". In principle, the formal theory is compatible with the existence of a disconnected domains of belief evolving in perfect independence of one another. But it is not really in the spirit of the view to think that this is a common occurrence. A typical belief web may contain tightly knit hubs of thematically connected beliefs, which are better integrated with one another than they are with beliefs outside the hub. But there will also be bridges linking the various islands, so that barely any beliefs are cleanly separated from the rest of the web. Thus, while it allows for more loosely connected beliefs than the classical picture does, the inquisitive view of belief states is still a thoroughly holistic vision.

2.6 Connected Views and Belief Updates

In §2.1, we observed that certain consequences of our beliefs come for free, so to speak: "beam-in-a-house" entailments are believed with no deductive reasoning required. The inquisitive theory accounts for this phenomenon in part by saying that beliefs are closed under parthood: if believing something necessarily involves believing its parts, then no cogitation is required to get to the parts. But the immediate consequences of a new belief are not always just its parts. Other downstream belief adjustments can be necessitated by the acquisition of a new belief. And a consequence that is obvious or immediate to one person need not be obvious to everyone.

The step from *Detroit has* ***more than 200*** *murders* to *Michigan has* ***more than 200*** *murders* may not be immediate to students at the University of Arizona, but it would be pretty obvious to anyone who grew up in Detroit. Had they run their experiment at the University of Michigan, Kahneman and Frederick would no doubt have found different results. Now that we understand the inquisitive model of belief a bit better, we are in a position to see how it is able to account for those cognitive differences. The way a belief is linked to your other views depends on its position in the web, and on fine-grained details about the content of your background beliefs. In particular, Michigan students' beliefs about Detroit and Michigan are in general more integrated than those of Arizona students.

To see how that works, let us revisit MITTEN STATE MURDERS I. Here Mandy brought her view on D, the number of murders in Detroit, to bear on the question M, the number of murders in Michigan. Specifically, she inferred from the fact that *there were **more than 200** murders in Detroit*, B^D, that *there were **more than 200** murders in Michigan*, A^M. (The inference really requires an additional premise, but to keep it simple, I will ignore the possibility that Detroit is in a different state.) There are various different ways such a transition could happen within the inquisitive theory, but here is one possibility. Prior to the inference, Mandy had two separate views, one about D, and the other about M; she did not have any view about the conjunctive question DM (that is to say, $DM \notin \mathscr{D}_{\mathbf{B}_{Mandy}}$). The deductive inference consisted in her *merging* these two views together into a single view about DM. After this merger, her views about D and M have both become part of a larger view and must therefore cohere. Consequently, the Detroit number can no longer exceed the Michigan number, and in particular the information that the Detroit number exceeds 200 is now brought to bear on M, resulting in the belief that the Michigan number is greater than 200 as well.

Let me go through this a bit more slowly. Consider the following answers:

$B_1 = \{ d \in D :$ according to d, there are over two hundred murders in Detroit $\}$

$B_2 = \{ dm \in DM :$ according to dm, there are over two hundred murders in Detroit $\}$

$A_1 = \{ m \in M :$ according to m, there are over two hundred murders in Michigan $\}$

$A_2 = \{ dm \in DM :$ according to dm, there are over two hundred murders in Michigan $\}$

B_1 and B_2 are truth-conditionally equivalent answers to D and DM respectively. They entail A_1 and A_2, which are truth-conditionally equivalent answers to M and DM respectively

Prior to the deductive inference, Mandy's view on D is B_1^D. For simplicity, let us say that her view on M is just the tautological M^M. She has no view on DM at all. We can think of the merger of the two views as the most conservative way of acquiring a view on DM while preserving her belief in B_1^D. For suppose Mandy believes B_1^D and also has a view on DM. Then she must at the very least believe the tautology DM^{DM}. By partial closure under conjunction it then follows she must also believe the conjunction of those two quizpositions, $B_1^D \wedge DM^{DM}$, which is equal to B_2^{DM}. Finally, A_1^M is part of B_2^{DM}, so by closure under parthood she believes A_1^M too.

This shows that Mandy can only acquire a view about DM by merging her views on D and M, bringing the answers to each question to bear on the other. The information could just as well go in the other direction. Suppose you start out with the view $\neg A_1^M$, that *there are* ***at most 200*** *murders in Michigan*, and a trivial view D^D about the number of murders in Detroit.[15] Then acquiring a view about the conjunctive question DM while preserving the view $\neg A_1^M$ requires you to adopt the view $\neg B_1^D$ as well, that there are ***at most 200*** *murders in Detroit*. If you start out with inconsistent beliefs about the two questions, for instance B_1^D and $\neg A_1^M$, then any merger will involve a choice about which of these beliefs to preserve.

One consequence of the merger-style treatment of this deductive inference is that as long as

[15] The ***negation*** of a quizposition A^Q, written $\neg A^Q$, is just the quizposition $\langle Q, Q \setminus A \rangle$.

Mandy hangs onto a view about DM, it will remain the case that she automatically brings any information about D to bear on M and vice versa. Thanks to the fact that these two views are now integrated into a bigger view, they no longer vary independently but are instead connected. And that seems plausible in this particular case. If, right after the conversation in MITTEN STATE MURDERS I, Mandy learns that there were actually 300 murders in Detroit, she will instantly infer that there were more than 300 in Michigan. And if she learns there were 400 murders in Michigan, she will instantly infer there were at most 400 in Detroit. By contrast, prior to this conversation her views were not yet linked; and while she could still have made those inferences, it would not have been instant or automatic.

A similar cognitive difference accounts for the fact that if you are from Detroit, you are less likely than others to forget about Detroit when it comes to assessing the number of murders in Michigan. If you are from Detroit, you probably have a lot of richly detailed beliefs about the city and also about Michigan, and these beliefs are likely closely intertwined, and integrated through bigger-picture questions like DM. As a consequence, any information about the one is liable to be brought to bear on the other instantly, since the relevant views are likely to be connected and well-integrated.

If two views are part of a single overarching view they must fully cohere. If two views are connected because they overlap with each other, they need only agree on the overlapping part. As indicated at the end of the last section, there are still further ways in which two views can be connected. As figure 5 illustrates, two views may overlap with a third view without overlapping one another (the quizpositions displayed on the lower tier are the shared parts of the quizpositions immediately above them).

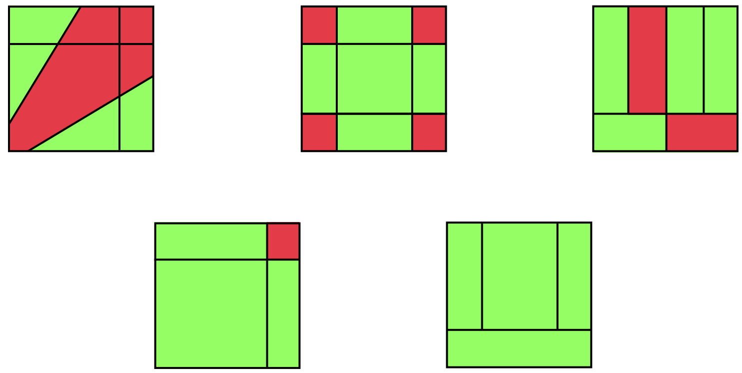


FIGURE 5: A DAISY CHAIN OF OVERLAPPING VIEWS

In a situation like this, a change in view on one end of the chain may be felt in every link of the chain. For instance, suppose some agent has the views depicted in figure 5. Then a change in view on the top left can necessitate a change in view on the top right, even though the two questions do not overlap, and even though they are not part of some larger overarching question. One such change is depicted in figure 6:

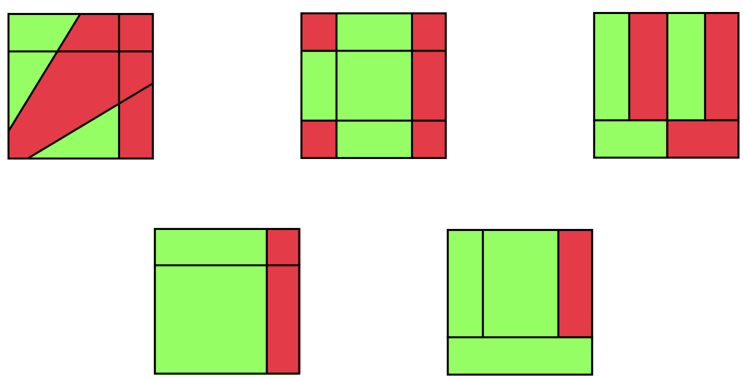


FIGURE 6: A BELIEF UPDATE ON LINKED VIEWS

Here, one additional cell of the top left question is ruled out. Doing so involves ruling out an additional cell in middle view too: this cell is in the overlap between the two. Consequently, a new cell in the overlap between the middle view and the view on the right is ruled out too. And thus the informational update on the left is automatically accompanied by an informational update the right because of the intercession of an appropriate background belief, which serves as a bridge between the two views. The inquisitive view makes very specific predictions about the circumstances under which views about different matters become linked in this way.

Let me give a concrete example. Suppose someone acquires a piece of information:

l: *Lincoln was president in 1862*

And suppose they have the following background information already:

f: *Lincoln served as president for four years*

a: *In 1870, either Lincoln or Grant was president of the United States*

Taken together, these three pieces of information f, a and l entail that

g: *In 1870, Grant was president of the United States*

According to the classical theory, all our beliefs are perfectly integrated, and the acquisition of the information l should cause the agent to believe g as well.

On the inquisitive theory, an agent may fail to see this consequence. The new information might be brought to bear on the situation in 1870 or it might not be. It depends on the questions involved. First suppose we take the three beliefs with truth conditions l, f and a to be L^Q, F^R and A^S respectively, where

Q: *Who was president in 1862?*

R: *How long was Lincoln president for?*

S: *Who was president in 1870?*

It is certainly possible for an agent to believe L^Q, F^R and A^S without believing G^S, that ***Grant*** *was* *president in 1870*. The question R does not overlap the other two questions, and so the belief F^{R} does not automatically serve as a conduit to connect the other two views. That seems right: surely it *is* possible to have each of these pieces of information without having made an inference to that conclusion.

However, a different but truth-conditionally equivalent belief would link the views Q and S in a way that does transmit the new information from one view to the other. Consider the question R', *When was Lincoln president?* Now let $\mathsf{F}'^{\mathsf{R}'}$ be another quizposition equivalent to f. The cells of R' represent possible start and end dates of Lincoln's presidency. F' is the set of all R'-cells at which there are four years between these two dates. Intuitively, you might think of the difference between believing F^{R} and $\mathsf{F}'^{\mathsf{R}'}$ in the following way: a view on R is just a number, or range of possible numbers, indicating the length of Lincoln's presidency. But to have a view on R', you need to mark the possible dates of Lincoln's presidency out on a mental timeline.

It is more cognitively demanding to have a view of the latter kind, but it pays off because a belief on R' bears on a greater set of issues. In particular, R' *does* overlap with Q: both contain the polar question *Was Lincoln president in 1862* as a part. R' also shares a part with S, namely *Was Lincoln president in 1870*. As a consequence, it turns out that if we keep the background beliefs A^{S} and $\mathsf{F}'^{\mathsf{R}'}$ fixed, adding the new belief L^{Q} *does* necessitate adding the belief G^{S}, that ***Grant*** *was president in 1870*.

Let's go through it step by step. The belief L^{Q}, ***Lincoln*** *was president in 1862* contains as a part an affirmative answer ***Yes*** to the polar question *Was Lincoln president in 1862*. By partial closure under conjunction, this part must be incorporated in the view on *When Lincoln was president*. Conjoining it with the existing view $F'^{R'}$ that *Lincoln was president **for four years***, you get *Lincoln was president **for four years between 1858 and 1866***. Now the answer ***No*** to the polar question *Was Lincoln president in 1870* is part of that quizposition. Using partial closure under conjunction again, this must become part of the view on S, *Who was president in 1870*. Conjoining it with the prior view on S, that ***Lincoln or Grant** was president in 1870*, we get the view G^S, that ***Grant** was president in 1870*.

Formally speaking, the two cases examined in this section can be modelled using the notion of an inquisitive *updates*. In the classical case, one can characterise informational updates as follows: for any information state **I** and intensional proposition p, **I** + p is the smallest classical information state containing **I** ∪ { p } as a subset. The notion of an inquisitive update is a generalisation of this idea:

> The ***inquisitive update*** of an inquisitive information state **I** by a quizposition A^Q, written **I** + A^Q, is the smallest inquisitive information state containing **I** ∪ { A^Q } as a subset.[16] (2.15)

[16] To see that inquisitive updates are always well-defined, note that (a) the set of all quizpositions is itself an information state (an incoherent one); and (b) any intersection of inquisitive information states is itself again an information state. It follows from (a) and (b) that for any set **S** of quizpositions, there is a smallest information state containing **S**. But just as classical updates do not always preserve the consistency of an information state, inquisitive updates do not always preserve *coherence*. So inquisitive believers can never update their belief state by a quizposition that fails to cohere with their prior views. Analogously, a classical believer can never update with information inconsistent with their prior beliefs. That is because such doxastic transitions require belief revision (see §2.8 for more on this).

Basically, the updated state $\mathbf{I} + \{A^Q\}$ represents the most conservative way the quizposition A^Q can be incorporated into the the inquisitive information state $\mathbf{I}$. It is maximally conservative in that it preserves all the quizpositions in $\mathbf{I}$, adding in as few quizpositions as needed to incorporate A^Q (while respecting closure under parthood and partial closure under conjunction). This notion of an update allows us to state our findings about MITTEN STATE MURDERS I and the Lincoln example more succinctly.

We found that whenever the anterior belief state $\mathbf{B}_X$ contains the tautologous quizposition DM^{DM}, which unifies $\mathbf{B}_X$'s views on D and M, then $\mathbf{B}_X + B_1^D$ must contain the quizposition that A_1^M, that *Michigan had* ***at least 200 murders***. And so part of the reason that Mandy failed to see this consequence of her belief B_1^D is that her views on D and M are not unified in this way. If one updates Mandy's antecedent belief state $\mathbf{B}_M$ with the missing tautology DM^{DM}, the resulting state $\mathbf{B}_M + DM^{DM}$ *does* contain A_1^M in addition to B_1^D. Thus certain deductive inferences can be captured as updates by tautologies — this will be the topic of §2.8 below. In the Lincoln example, we found that, if an anterior belief state $\mathbf{B}_Y$ contains the premises $F'^{R'}$ and A^S, the updated state $\mathbf{B}_Y + L^Q$ is guaranteed contain G^S. But if $\mathbf{B}_Y$ only contains the intentionally equivalent premises F^R and A^S, the posterior state $\mathbf{B}_Y + L^Q$ is not guaranteed to contain G^S. Again, this accounts for the contrast between those for who, upon learning l, draw the conclusion g right away, and those for whom the inference is not immediate.

In both cases, the inquisitive theory of belief explains how the upshot of a new piece of information depends, not only on an agent's background information, but also on the way an agent's views are linked together by their background beliefs. As both cases show, beliefs about bigger questions make for more connections and result in greater overall coherence of our beliefs. The biggest question of all is the discrete question $\mathsf{W} = \{ \{w\} : w \in \mathscr{W} \}$ (also known as the Big Question; e.g. Roberts 2012). A believer with a view about W would have beliefs that are fully closed under entailment and conjunction, which is to say they would be a classical believer. The reason that ordinary, finite agents cannot achieve this ideal is that there is a cost associated with keeping track of the big questions: it is cognitively demanding to do so.

2.7 Computational Limitations

Some inquisitive updates will radically alter a belief state, transforming views throughout the agent's web of belief. Others affect only the view concerned and some of its parts. Taking inquisitive updates as models of real-life belief acquisitions, this is in accordance with experience: some of our discoveries are superficial, and can be incorporated without affecting the one's other views much at all. But others affect deeply held beliefs, and cannot be fully accepted without sweeping, comprehensive changes in view on many matters. The natural expectation is that changes in view of the latter kind are more cognitively demanding.

The immediate consequences of a new belief only come for "free" in the sense that no cogitation *beyond* the acquisition of the belief is needed in order to believe those consequences. The flip-side of that is that the more of those consequences there are, the more complex and pervasive a cognitive change is needed for the belief to be acquired. Thus it is a non-trivial cognitive achievement to make one's beliefs well-connected.

Let me illustrate this with an example. I have a terrible sense of direction. When I am new in a city, I will figure out how to get from the hotel to the central square, say, and how I can get from the central square to the museum or the river bank, and from the museum to the restaurant. But having gathered all this information, I still will not be able to work out an even a half-way efficient route back from the restaurant to the hotel. Without a map, the safe option is just to retrace my steps: return to the museum, then back to the central square, then to the hotel; else I will probably get lost. My friend is very different. Given just the same information, she will identify the shortest way back in a heartbeat, even if it runs through a neighbourhood she has never seen before.

Maybe my friend just has a better memory for these things than I do. But what is more important, I think, is the way she puts it together. My mental geography of the city is a chaotic patchwork of partially overlapping little maps, patched up with landmarks, mnemonics and other crutches. I only have answers to small, local questions about the city's geography, which are barely connected to one another, and none of which have any bearing on unexplored areas of the city. My friend, on the other hand, sees the bigger picture. Her views about the geography of the city are more robustly connected because she keeps track of the big picture: she has views about the overall layout of the city that integrate her detailed views about the parts. This puts her in a position to see we have walked in a big circle, and that the hotel is just a few blocks away. Her beliefs answer bigger questions, and are better linked together.

Maintaining a high level of integration between geographical beliefs is cognitively demanding. My friend is better at it than I am, and you can improve your skills with practice. London cab drivers make for an extreme example. To get a taxi license, they must pass a harrowing exam called *The Knowledge*, requiring three to four years of intensive training. Over the course of this period, they acquire the ability to efficiently deploy a vast amount of detailed geographical information in order to determine the fastest route from one place in London to another. It has been shown that in successful trainees, this learning process results in a significantly enlarged posterior hippocampus (Maguire et al. 2000, 2006, 2011).

On the one hand, this is a remarkable example of the way neuroplasticity allows human beings to go beyond their innate cognitive abilities. On the other hand, the finding that additional grey matter is required to achieve such levels of skill clearly implies that there are limitations to those abilities, and limitations to how far they can be stretched. There is only so much grey matter in a human skull, and there is only so much new grey matter one can grow — if only because of space constraints. Already in the case of the cab drivers, the growth leads to a bit of a squeeze: the growth of the posterior hippocampus had a negative long-term effect on the anterior hippocampus (Maguire et al. 2000, 2006).

In §2.1 we discussed how, according to the classical theory, an agent who knows the exact scrambled state of a Rubik's cube would also know how to unscramble it. In the inquisitive theory, we can model the mental state of someone who has all this information without knowing how to unscramble the cube. For instance, they might have the configuration of each side memorised, and remember how to fit the sides together. This knowledge would give them the ability reproduce the configuration of the entire cube, but it would not give them the ability to solve the cube. For instance, it would not tell them which side to twist first.

What if an inquisitive agent were to put all this information together into an answer to the giant question *What is the configuration of the Rubik's cube*, which has at least forty-three quintillion cells? Well, then they would be just like the classical agent, and they would just *see* how to solve the cube: for since the configuration of the cube determines the sequence of steps needed to solve it, the question *How to solve the cube* is part of this enormous question. But given our actual computational limitations, this is a merely theoretical possibility: we human beings are unable to achieve that level of fluency with Rubik's cubes. (Some people can solve Rubik's cubes very quickly. But they do it by memorising algorithms, and not by answering a forty-three quintillion cell question.[17])

Note this is not just about failing to put all the information together into a single belief. For instance, an agent who can reliably tell whether or not a given configuration is identical to that of the cube under consideration knows that *The cube is in configuration X* in answer to the polar question *Is the cube in configuration X or not*. So such an agent has put all the required information together into one belief. There is nothing unrealistic about supposing someone might believe this quizposition, and it does not imply an ability to unscramble the cube. Our computational limitations put a ceiling on the size of the questions we mere mortals can have answers to — not on the logical strength of the answers.

How high is this ceiling, exactly? How big can a question get before it becomes hopelessly intractable for an adult human being of average intelligence? The abstract theory does not tell

[17] The most popular speed solving method in use is the Fridrich method, designed by steganographer Jessica Fridrich — see www.ws.binghamton.edu/fridrich/cube.html.

us. Nor should it: just like classical view of belief and action, the inquisitive view is supposed to apply not only to human beings, but to a wide range of real and possible agents, which may all have different computational limitations. The theory allows us to pose questions about our cognitive computational limits more precisely. But to find the answers to such questions, empirical work is needed.

As a matter of fact, there is already some work in psychology that seems to bear on these questions. In particular, Philip Johnson-Laird's (1983, 2006) work on mental models is relevant, as is Philipp Koralus and Salvador Mascarenhas's (2013, 2018) work on the erotetic theory of reasoning. These authors have conducted experiments that provide empirical support for the basic idea that it is cognitively demanding to keep track of large questions with many alternatives. Johnson-Laird actually puts a number on it, estimating that an ordinary human being can keep track of no more than five to seven alternatives at a time in working memory. Given the particularities of Johnson-Laird's theory, this observation does not straightforwardly carry over to the present context. I would estimate that we can have views about questions with much more than seven cells, particularly for beliefs on familiar topics about which we can make very quick inferences. But it is safe to assume that the limit is nonetheless a good deal lower than forty-three quintillion, which is what we would need to become instant Rubik's experts.

One more remark about the Rubik's cube example. Given chapter 1, it may come as a surprise that I should be hesitant to attribute knowledge of *What the configuration of the Rubik's cube is* in this case, given that the agent is able to reproduce that configuration. Isn't this the ability that corresponds to knowing *What the configuration of the Rubik's cube is*? Well, it is, in the sense that having this knowledge would entail having the ability to reproduce the configuration. But the entailment does not go the other way. We can also account for the agent's ability to reproduce the configuration of the cube as a consequence of a collection of simpler abilities, each one requiring more modest, less demanding beliefs. For instance, say the agent's beliefs about each of the six sides allow them to reproduce the configuration of that side. Having explained how the agent was able to reproduce the configuration of each side of the cube, we have also explained how they could reproduce the configuration of the entire cube. In this manner, we can explain the agent's abilities on the basis of their beliefs, while at the same time avoiding the attribution of unrealistic levels of deductive closure (see §3.4 for more on this general strategy).

2.8 Tautological Updates and Deductive Reasoning

One of the upshots from the last section is that it is an inevitable part of the human condition that we can never see all the entailments of our beliefs. Still, we can extend our logical reach considerably through mental effort and deductive reasoning. In particular, human beings have an ability to pose new questions, thereby forging new connections in our web of beliefs and allowing us to see consequences of our beliefs that were hidden at first.

As suggested by the discussion of MITTEN STATE MURDERS I in §2.6, one useful application fo the notion of an inquisitive update is to model this process, yielding a simple and attractive way to model some instances deductive reasoning. Once we understand why an agent was failing to see a certain consequence of what they believe, and which connections they were missing, this yields a natural hypothesis about what was added in the deductive step from their prior belief to seeing this new consequence.

Take the case of Mandy in MITTEN STATE MURDERS I, who believes B_1^D, *Detroit had **at least 200 murders***, but not A_1^M, *Michigan had **at least 200 murders***. As shown in §2.6, since her belief state $\mathbf{B}_m$ contains B_1^D but not its consequence A_1^M, $\mathbf{B}_m$ must be lacking the tautological quizposition DM^{DM}. What is more, updating with this necessary truth leads to a posterior belief state $\mathbf{B}_m + DM^{DM}$ that does contain A_1^M. And thus one natural way to account for Mandy's cognitive achievement at the end of MITTEN STATE MURDERS I is to say that she performed the update $\mathbf{B} \mapsto \mathbf{B} + DM^{DM}$.

A similar observation can be made about the Lincoln example from §2.6. Suppose $\mathbf{B}_l$ is an inquisitive belief state that contains F^R, *Lincoln served as president **for four years***, A^S, ***Lincoln or Grant** was president in 1870* and L^Q, ***Lincoln** was president in 1862*. And suppose it is missing the conclusion G^S, ***Grant** was president in 1870*. Then $\mathbf{B}_l$ cannot contain the tautological quizposition $R'^{R'}$ — i.e. the trivial answer to the question *When was Lincoln president*. And the state $\mathbf{B}_l + R'^{R'}$, the result of incorporating that tautology, *does* contain G^S. Thus, we can again model the deductive step from F^R, A^S, and L^Q to G^S as an instance of the update by a tautological quizposition.

Let us refer to updates of the form $\mathbf{I} \mapsto \mathbf{I} + Q^Q$ as ***tautological updates***. As the Mandy and Lincoln examples illustrate, it is natural to use such updates to model certain simple deductive inferences. Not all deductive inferences can be understood as the result of tautological belief updates. For instance, hypothetical reasoning proceeds on the basis of a body of information that includes tentative hypotheses and not just the agent's established beliefs (as in *reductio ad absurdum* and in proof by cases). Nonetheless, tautological updates make for a neat and intuitive example of the way the inquisitive picture, having eliminated closure under entailment, makes room for systematic theorising about deductive inference.

Since tautological quizpositions do not carry any new information, the beliefs an agent acquires as a result of a tautological update are always beliefs in quizpositions that were already entailed by the agent's antecedent beliefs. Nonetheless, tautological updates can lead to new, contingent beliefs, and they can percolate through an agent's web on belief the same way as contingent updates do. For instance, it is easy to imagine how, given suitable background beliefs, Mandy's finding that *Michigan had* ***over two hundred*** *murders* has a downstream effect on her other views about Michigan.

Figure 7 illustrates the abstract situation. Here we start out with a belief state containing views on Q_0, R and S, and then perform a tautological update that refines the question Q_0 to Q_1. The resulting view $A_1^{Q_1}$ has the same truth conditions as $A_0^{Q_0}$. However, the new question Q_1 this view addresses overlaps with R, and $A_1^{Q_1}$ rules out some cells in the overlapping part. Thus the update strengthens the agent's view on R from B_0^R to B_1^R. This change in view about R in turn affects the agent's view on the question S, which also overlaps with R. Thus the update by $Q_1^{Q_1}$ causes a change in view about S, even though Q_1 does not overlap with S at all. In the same way, the update can percolate further down the daisy chain, spreading throughout the agent's web of questions.

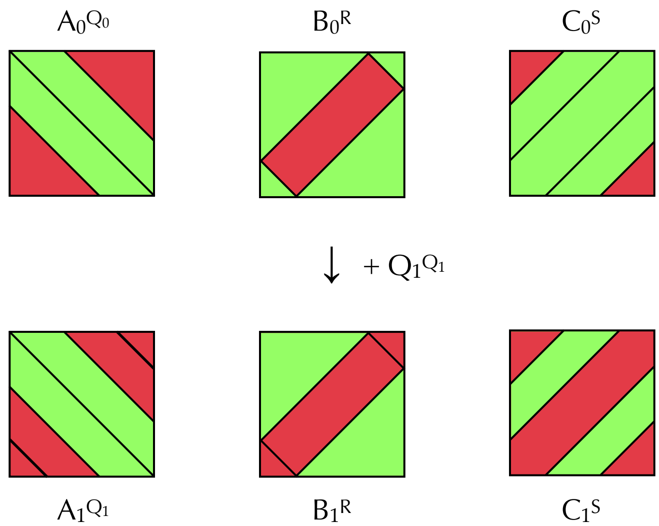


FIGURE 7: A TAUTOLOGICAL UPDATE ON LINKED VIEWS

Thus, on the inquisitive model, the acquisition of new tautologous beliefs can lead to all sorts of new contingent beliefs, including beliefs about questions unrelated to the question the added tautology is about. We can think of a tautological update as modelling what happens when an agent poses a new question for the first time, where we understand the "posing" of a question Q as the acquisition of an inquisitive view about Q. The mapping $\mathbf{I} \mapsto \mathbf{I} + Q^Q$ is the natural formalisation of this idea: its output $\mathbf{I} + Q^Q$ is the smallest extension of the information state $\mathbf{I}$ that includes a view about Q.

Thus the inquisitive model of belief naturally yields a formal implementation of the idea of deduction as a question-guided endeavour — compare Pérez Carballo 2016, and Friedman 2017, 2019. Think of the slave boy from Plato's *Meno*. Guided by the questions that Socrates asks him, the boy reasons his way to the conclusion that the diagonal of a square of size one is equal to the side of a square of size two. From the outset, the boy already has all the basic geometric intuitions he needs to figure this out. But he only arrives at the right conclusion after thinking through Socrates' strategically posed questions.

The first half of this chapter (§2.1-5) concerned the *statics* of belief. With the introduction of the notion of an inquisitive belief update, we have now made our first foray into the dynamics of inquisitive belief. Belief updates model the *addition* of new beliefs to a state, with preservation of the agent's prior beliefs. There are many other kinds of changes a belief state can undergo. A fuller dynamics of belief may include a treatment of other kinds of belief change as well. In particular, in both the classical and inquisitive theory, simple belief updates cannot account for the situation where agents acquire new information that conflicts with their extant beliefs: in those situations, the agent is forced to give up some of their old beliefs, in addition to acquiring a new one.

In the classical theory, this issue comes up because classical updates do not preserve consistency: if you update a classical information state **I** with an intensional proposition p that is inconsistent with it, you always end up with the inconsistent classical information state, which contains every intensional proposition, and does not represent a possible classical belief state. In the inquisitive theory, inconsistency does not by itself present problems, since there are possible inconsistent belief states. Instead, belief revision is forced when an update fails to preserve the *coherence* of the information state. This can happen even if the agent updates with a tautology. That situation can arise when an agent comes to see the inconsistency of their previous beliefs by means of deductive reasoning.

Consider for example MICHIGAN MURDERS II. At the end of this version of the story, Mandy's belief state $\mathbf{B}_i$ is inconsistent, since it includes both the quizposition B_1^D, that *Detroit had* ***over 200*** *murders*, and also the quizposition $\neg A_1^M$, that *Michigan had* ***no more than 200*** *murders*. As discussed, $\mathbf{B}_i$ is an inconsistent but coherent belief state. However, if Mandy were to combine her beliefs about Detroit and Michigan in this situation, she would run into incoherence: the updated information state $\mathbf{B}_i + DM^{DM}$ includes both the quizposition A_1^M and the quizposition $\neg A_1^M$, and therefore also their conjunction $\perp^M$.

Clearly, this update does not accurately represent Mandy's change in doxastic state: the incoherent information state $\mathbf{B}_i + DM^{DM}$ is not even a possible belief state. In particular, Mandy would not end up believing $\perp^M$, that *Michigan had* ***no more than and more than 200*** *murders*. This is a situation in which it becomes evident to Mandy that her prior beliefs were inconsistent, not a situation in which her beliefs themselves turn incoherent. How does that realisation in fact affect Mandy's beliefs? Well, if Mandy is to acquire the new belief DM^{DM}, she must abandon some of her prior beliefs to avoid incoherence. In the case as presented in §2.4, Mandy would probably discard her prior belief that $\neg A_1^M$, recognising that her grounds for believing A_1^M are stronger. In a different scenario, Mandy might give up B_1^D and A_1^M instead, or suspend judgment on the matter.

The reason this doxastic transition cannot be modelled using updates alone is that it inevitably involves Mandy discarding one of her prior beliefs. Notoriously, modelling belief subtraction or revision is a much trickier matter than modelling simple belief addition. There are, however, plenty of proposals in the literature — notably Alchourrón, Gärdenfors and Makinson 1985; Van

Ditmarsch, Van der Hoek, and Kooi 2007, Van Benthem and Pacuit 2011. While I will make no attempt to do so here, I think considerable interest attaches to the project of extending such models of belief revision to the inquisitive setting, because the additional resources of the inquisitive framework may put us in a position to do better justice to the relevant cognitive phenomena. In particular, I think the weblike structure of inquisitive beliefs may help us understand why belief revisions ordinarily have fairly localised effects.

2.9 Conceptual Limitations

Not all tautological belief updates can be achieved through deductive reasoning. In §2.7, we already saw that our finite computational resources limit the size of the questions we can hope to acquire views about. But our conceptual resources are limited too, and this puts a different kind of constraint on the questions we can ask. For example, consider the following question:

P: *Are pineapples from South America?*

It is certainly not beyond the capacity of the human mind to form a view about the simple polar question P — many of us have. And yet, we can be certain that Aristotle never did it.

In the time of Aristotle, no European had ever been to South America, or seen a pineapple. It is not for lack of computational power that Aristotle was unable to ask this question. Deductive reasoning alone could never have gotten Aristotle from his actual belief state $\mathbf{B}_A$ to the state $\mathbf{B}_A + \mathsf{P}^{\mathsf{P}}$. To make this transition, Aristotle would have had to learn some new concepts first, in particular the concept of a pineapple and the concept of South America. Acquiring those concepts, in turn, would have required pineapples to be discovered two millennia earlier, and it would plausibly require some world knowledge on Aristotle's part. He could not have arrived at $\mathbf{B}_A + \mathsf{P}^\mathsf{P}$ from the armchair.

Likewise, before you started reading this paragraph, you (the reader) were in no position to reason yourself to the belief that *Either the village of Komminenivaripalem is in the mountains or it is not in the mountains*, for the simple reason that you had never heard of that village. (In a way, I ruined this particular example by telling you about Komminenivaripalem. Now that you know about the place, you are plausibly in a position to start forming beliefs about it. But an hour ago you could not. And even right now I could make an analogous point about a different village — I am not telling you which one!)

Here is a slightly more complex example from Stalnaker. King William believed he could avoid war with France. But he did not, it seems, believe that he could avoid *nuclear* war with France (Stalnaker 1984; Yalcin 2011, 2018 also discuss this example, making a similar point about it). It's not because of a lack of logical acumen that this 17th-century monarch failed to see that particular consequence of his beliefs. William had no idea what a nuclear weapon was, and it seems no amount of reflection or deduction on his part would have fixed that. And yet he is only a tautological update away from seeing this consequence.

To see this in a bit more detail, let E be the polar question *Will England have a war with France or not*. Now suppose King William's belief state $\mathbf{B}_W$ contains the quizposition A^E, that *England will avoid war with France*. Let F be the tripartite question *Will England have a nuclear war with France* *or some other kind of war or no war at all*: this question contains E. The tautologous answer to F is F^{F}, that *Either England will have a nuclear war with France or some other kind of war or no war at all*. If William were able to distinguish the possibility of nuclear war from other kinds of war, he could perform a tautological update, moving from the prior state $\mathbf{B}_W$ to the state $\mathbf{B}_W + \mathsf{F}^{\mathsf{F}}$. Besides the quizposition A^{E}, William's posterior state $\mathbf{B}_M + \mathsf{F}^{\mathsf{F}}$ would then include a view on the new question F that includes the quizposition B^{F}, *England will avoid nuclear war with France*. Again, reason William is in no position to make this transition is not because he does not have the computational resources: rather, he lacks the relevant concepts.

In §5.7 below, I suggest a different, more contentious example of a tautological belief update that cannot be achieved a priori: the famous case of Lois Lane. Lois believes that *"Superman" refers to Superman*, and she also believes that *"Clark Kent" refers to Clark Kent*. At all worlds where those two beliefs beliefs are true, it is also true that *the names "Superman" and "Clark Kent" refer to the same person*. Thus Lois' discovery that this is so can be modelled as a tautological belief update. Her belief about S, *To whom the name "Superman" refers*, and her belief about K, *To whom the name "Clark Kent" refers* can be brought together by joining them into a view about SK, *To whom do the names "Superman" and "Clark Kent" refer*. But clearly this is a substantial discovery, not an a priori finding. Thus the update $\mathbf{B}_L \mapsto \mathbf{B}_L + \mathsf{SK}^{\mathsf{SK}}$ is not a deductive step for Lois: making this step would require her to sort out her identity confusion first (for more details on this, see §5.7).

2.10 A Good Question is Hard to Find

Besides our computational and conceptual limitations, another obstacle in the way of seeing the consequences of one's beliefs is knowing which questions to ask. Because any tautological update requires a certain amount of cognitive effort to process, and because resources are limited, we cannot form a view on *every* question. So when we engage in deductive reasoning, we must inevitably make choices about which questions to look into (cf. also §3.4 below). In the *Meno*, Socrates' strategic questioning helps the slave-boy precisely because it relieves him of this part of the cognitive labour. As that example illustrates, sometimes the most difficult part of a deduction is not performing the update itself, but rather knowing which update to perform. Hitting on the right question to ask can require insight or luck.

Here is a case from the psychology literature which illustrates the point (Levesque 1986, Toplak and Stanovich 2002). Based on the following three pieces of information, can you say whether or not an unmarried person is looking at a married person?

1) Jack is looking at Kate and Kate is looking at George
2) Jack is unmarried
3) George is married

Take a moment to picture the situation and think it through.

In Toplak and Stanovich's survey, 86% of the respondents answered that the correct answer cannot be determined on the basis of the information provided. However, as a matter of the fact it can. This becomes easy to see once you are given the following hint:

4) Either Kate is married or she is unmarried

Once those two possibilities are separated, the answer becomes clear. If Kate is married, then Jack is an unmarried person looking at a married person, because Jack is looking at Kate. If Kate is unmarried, then she herself is an unmarried person looking at a married person, because she is looking at George. It is striking how an instance of the law of the excluded middle transforms an otherwise elusive inference into a no-brainer. This makes (4) an unusually simple and elegant example of an informative tautology.

The inquisitive view of belief accounts for this phenomenon as follows. Conjoining the the three given premises (1-3) is insufficient to arrive at the conclusion that an unmarried person is looking at a married person. But it *is* part of the conjunction of (1-4). To see this, associate premises (1), (2) and (3) with the quizpositions A^L, B^J and C^G respectively where

L: *Out of Jack, Kate and George, who is looking at whom?*

J: *Is Jack married?*

G: *Is George married?*

The task confronts subjects with something like the following question:

Q: *Who out of Jack, Kate and George is unmarried? And who is looking at a married person?*

And the target conclusion is:

D^Q: *One of Jack, Kate or George is an unmarried person looking at a married person*.

In approaching this problem, the natural first step is to put all the given information together into a single representation. This may be modelled as an update with the tautologous quizposition LJG^{LJG}. This takes a state in which A^L, B^J and C^G are believed individually to a state where their conjunction ABC^{LJG} is also believed. However, this update is not sufficient. The conjunction ABC^{LJG} entails the conclusion D^{Q}, but because Q is not part of LJG, the quizposition ABC^{LJG} does not *contain* D^{Q} as a part. A further step is required to get from the view ABC^{LJG} to the target conclusion D^{Q}.

Some other simple strategies to get at the answer also fail in this case. For instance, it is natural to break up Q into simpler questions: *Is Jack an unmarried person looking at a married person*, *Is Kate an unmarried person looking at a married person* and *Is George an unmarried person looking at a married person*. ABC^{LJG} fails to settle the first two questions, and entails a negative answer to the latter. From these results one might reasonably conclude that the given information is insufficient to settle whether D^{Q} is true. Plausibly, that is where the inquiry halts for most of Toplak and Stanovich's respondents.

The only way to arrive at the target conclusion is to make a further distinction. We need to separate two possibilities that LJG joins together: namely the possibility that *Jack and Kate are unmarried and only George is married*, and the possibility that *only Jack is unmarried and Kate and George are married*. Separating of these two scenarios involves conjoining ABC^{LJG} with the content of (4), *Either Kate is married or not*. This is the tautologous quizposition K^{K} where

K: *Is Kate married?*

After that further adjunction, the subject's overall view of the situation is $ABCK^{LJGK}$; and since LJGK *does* contain Q as a part, it follows that they now believe the target conclusion D^{Q} as well.

There is some amount cognitive effort involved in making the extra distinction taking you from

ABC^{LJG} to $ABCK^{LJGK}$. But the students Toplak and Stanovich interviewed could all have made this further reasoning step if prompted. The explanation for why most of them failed to take this step does not lie in the intrinsic difficulty of the update. Rather, the students must have *overlooked* the question K for some reason. It may be that it simply did not occur to them: there is experimental evidence that reasoners are in general better at *recognising* a good question when it is presented to them than they are at coming up with good questions on their own (Rothe, Lake and Gureckis 2018). It is also also possible they did consider the question, but made an *a priori*, metacognitive judgment that it was not worth looking into.

In this particular context, two factors may contribute to that decision. Firstly, the fact that they were not given information about Kate's marital state could be taken as an indication that this issue is irrelevant. There is empirical evidence showing that reasoners are generally reluctant to think through a question when they know in advance that the answer is unknown (Toplak and Stanovich 2002, Shafir 1994, Tversky and Shafir 1992). Secondly, the fact that other lines of inquiry do not resolve the matter may have given rise to an overriding impression that the information provided was inadequate, and that further cogitation is pointless.

The fact that we cannot ask every question means we inevitably need some heuristic that will decide which questions to ignore and which to consider. And that heuristic must be prior to actually thinking through the questions. For instance, conjoining ABC^{LJG} with the tautology *Either George owns two camels or he does not* also make some new consequences available. But since you know *a priori* that none of those consequences will be of help in resolving the task at hand, you would never look into that question. It is safely ignored. In the case at hand, the question K apparently does not appear fruitful to most people, although that appearance is misleading.

2.11 Derivation of the Inquisitive View

Given that the notion of a propositional part comes out of the literature on truthmakers, and was developed for independent reasons, it is really remarkable that it should re-emerge so very naturally from the inquisitive way of thinking about belief-guided action. In this subsection I give a formal demonstration of the connection between the inquisitive view of individual beliefs (1.12) and of belief states (2.12) that was established informally in §2.5. Each of the three inquisitive coherence constraints is a strict weakening of a classical constraint: closure under parthood is weaker than closure under entailment, partial closure under conjunction is weaker than full closure under conjunction, and coherence is weaker than consistency. (Recall that the definition for entailment between quizpositions is the same as the definition for entailment between intensional propositions: one quizposition A^Q *entails* another quizposition B^R just in case B^R is true at all worlds where A^Q is true.) The derivation of each requirement closely parallels the classical one from §2.2. In particular, the derivations here also rely on something like the Quacks-Like-A-Duck principle (2.3).

The starting point, recall, is the inquisitive view of individual beliefs (1.2), formalised in (1.12):

> **Individual Beliefs (Inquisitive Formalisation)**. Belief is a relation between agents and quizpositions. An agent believing the quizposition A^Q is disposed to avoid A-dominated options in decision situations that raise Q. (1.12)

The reason an inquisitive agent cannot believe a contradiction is simple: (1.12) entails that an agent believing $\perp^Q$ would need to be disposed to perform *every* option in decision situations that raises Q, which is absurd. That establishes the coherence requirement.

Next up is closure under parthood. Let B^R be any part of A^Q, and suppose some agent α believes A^Q. We want to show that α also has the disposition (1.12) associates with believing B^R, namely to avoid B-dominated options when facing R. So suppose α makes a choice Δ that raises R, and suppose some option **a** in Δ B-dominates some alternative **b** in Δ. Then by (2.13), Δ also raises Q, and since A^Q entails B^R, it follows from the fact that $\mathbf{a}(r) > \mathbf{b}(r)$ for all $r \in B$ that $\mathbf{a}(q) > \mathbf{b}(q)$ for all $q \in A$. And thus, because α avoids A-dominated actions when faced with Q, α will forego the option **b**. So α will forego every B-dominated option when confronted with R. And in general, α will have every disposition that (1.12) associates with believing any part of A^Q. Thus (1.12) motivates closure under parthood. This argument does not generalise to show that inquisitive beliefs are closed under entailment. The reason is that, given (1.12), a stronger belief is guaranteed to address the same as choices a weaker belief only if it answers a bigger question.

That leaves partial closure under conjunction. As in the classical case, we need to appeal to composite choices here. Suppose A and B are answers to Q and R respectively, where R is any part of Q. We can again show that no agent who fails to avoid AB-dominated options when faced with Q can succeed in both avoiding A-dominated options when faced with Q and also B-dominated options when faced with R. Consequently, anyone who believes A^Q and also believes B^R, and hence succeeds on both fronts, also has the disposition (1.10) associates with believing AB^Q.[18]

This proof considers a composite choice consisting of a Q-raising choice Δ^Q and an R-raising choice Δ^R, and requires that the belief in one of these conjuncts, say A^Q, guides not only the agent's choice in Δ^Q but also the composite choice. That is to say, the argument only goes through if Q addresses the composite choice as well. However, the combined outcomes of two choices Δ^Q and Δ^R in general depends *both* on the question Q and also on the question R: that is to say, the composite decision problem raises the question QR. And QR = Q just in case R is part of Q. That is why we do not get full closure under conjunction, but only partial closure.

There is one more question left to address. We showed that coherence can be established by the same argument used in the classical case: incoherent beliefs would yield impossible predictions about what the agent will do. But why doesn't this reasoning imply cross-question consistency as well? After all, a single decision problem can raise multiple different questions. The same

---

[18] Let A and B be answers to Q and R respectively, where R is any part of Q. Suppose some agent α is disposed to avoid B-dominated options when faced with R, but *not* to avoid AB-dominated options when faced with Q. So α sometimes chooses **b** over the alternative **a** in a Q-raising choice Δ where $\mathbf{b}(q) < \mathbf{a}(q)$ for all $q \in AB$. We show that having made such a choice, α cannot in general avoid A-dominated options when faced with Q either.

Suppose that right after choosing **b** over **a**, α faces a choice $\{ \mathbf{o}, \mathbf{x} \}$ between a constant option **o** yielding 0 utility, and an option **x** that returns a positive value ε if B^R is true, but imposes a cost $-C$ if B^R is false. Since R addresses this choice, and **x** strictly B-dominates **o**, the agent will choose **x** over **o**. Now we can select ε and C so that

$$\mathbf{b}(q) + \mathbf{x}(q) \quad = \quad \begin{cases} \mathbf{b}(q) + \varepsilon \ < \ \mathbf{a}(q) & \text{for all } q \in AB \\ \mathbf{b}(q) - C \ < \ \mathbf{a}(q) & \text{for all } q \in A \setminus AB \end{cases}$$

For instance, put $\varepsilon = \frac{1}{2} \cdot \min\{ \mathbf{a}(q) - \mathbf{b}(q) : q \in AB \}$, and $C = \max\{ \mathbf{b}(q) - \mathbf{a}(q) : q \in A \setminus AB \} + 1$. But then α's overall course of action $\mathbf{b} + \mathbf{x}$ is A-dominated by $\mathbf{a} + \mathbf{o}$. What is more, the composite decision problem raises QR, and because R is part of Q, QR = Q. So α chose an A-dominated option while faced with Q.

reasoning implies that the agent's beliefs about all the questions raised need to render the consistent verdicts. But that seems to imply that the agent's beliefs about those different questions should be consistent with one another as well.

It is true that the argument takes us a little bit further, but a more careful analysis shows it does not take us all the way to consistency (and the condition it does justify is already implied by the three conditions established above). To see this, suppose some decision problem Δ raises both Q and R, and suppose the agent has views V^Q and W^R about these questions. Then, in order to guarantee consistent predictions, we need it to be the case that V^Q and W^R together do not render every option in Δ strictly dominated. This implies that the views V^Q and W^R have to harmonise in a way, but not that they have to be consistent with one another.

To see this, let me begin by noting that it follows from the fact that Q and R both address Δ that the overlap of Q and R, call it S, must also address Δ. To see this, let **a** be any option in Δ: we need to show that for any worlds w and v such that $w \sim_S v$, $\mathbf{a}(w) = \mathbf{a}(v)$ (definition 1.8). To see this let $p_w = \{ v : \mathbf{a}(v) = \mathbf{a}(w) \}$. Since Q and R both address Δ, it must be that p_w is both equivalent to some (possibly partial) Q-answer and also to some (possibly partial) R-answer. But then p_w must be equivalent to a (possibly partial) S-answer, for if there was an S-cell bigger than p_w, S would not be the largest common part of Q and R. Hence it follows from the fact that $v \sim_S w$ that $v \in p_w$, whence $\mathbf{a}(w) = \mathbf{a}(v)$.

Now because Q and S both address Δ, and S is part of Q, all and only the actions in Δ that are V-dominated are also V/ S-dominated, where V^Q is the agent's view on Q. For the worlds w at which V/S is true all share an S-cell with a world where V^Q is true, and the utility value of the options in Δ does not vary within an S-cell. For the same reason, letting W^R be the agent's view on R, the actions in Δ that are W-dominated are also W/S-dominated. Thus the requirement that there be some action in Δ that is neither V-dominated nor W-dominated boils down to the requirement that there is an action in Δ that is neither V/S-dominated nor W/S-dominated. And this is guaranteed just by the fact that the S-parts V/S and W/S of the agent's view be consistent with each other, which is a weaker constraint that the consistency of V^Q and W^R.

In short: to guarantee that the views V^Q and W^R give the same guidance in all decision situations where both of them apply, we do not need the views to be consistent. We only need them to be consistent in the overlap. To see that this is indeed a weaker requirement, consider figure 8 below. The two views displayed at the top are inconsistent with one another: there is no possible world where both are true. They also overlap with one another: the overlapping part is displayed underneath. But in spite of their overall inconsistency, what these two views say about the question in the overlap is consistent, and this is enough to guarantee that they will provide consistent guidance in any decision situation addressed by both of the big questions.

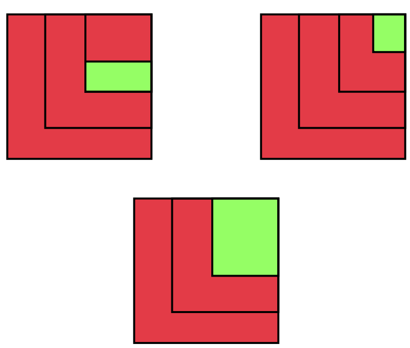


FIGURE 8: INCONSISTENCY BETWEEN OVERLAPPING VIEWS

Now, you may recall that closure under parthood and partial closure under conjunction guarantee that V/S and W/ S are not only consistent with one another but in fact identical, as is in fact the case with the views represented in figure 8 (this is the principle (2.14) above). For that reason, this requirement does not need to be added to the characterisation of an inquisitive belief state as a separate condition.

Say a set of quizpositions **S** is *unbelievable* if and only if, given (1.12), no coherent set of behavioural dispositions is described by the beliefs in **S**. And say **S** is *incomplete* if and only if, given (1.12), some strict superset of **S** describes the exact same behavioural dispositions as **S** does. Then we can state the result of this section as follows: the coherence constraints of (2.12) rule out all and only those belief states that are either unbelievable or incomplete.

Chapter 3.
Inquisitive Decision Theory

In this chapter, we add the final building block we need for a fully fledged inquisitive decision theory: the account of uncertainty. To bring uncertainty into the picture, we simply replace the *views* from the classical and inquisitive treatments of belief states with probability distributions. All the credences of a classical agent form a single probability distribution, while the credences of an inquisitive agent make a probability distribution for each question on which they have a view. Those question-directed probability distributions are not separate. They are linked by their thematic connections in much the way full inquisitive beliefs are linked: where they concern overlapping questions, the inquisitive credences match on the overlapping part.

Sections §3.1-2 discuss the classical treatment of uncertainty, and explain how the limitations of the classical theory observed above emerge in a probabilistic context. §3.3 introduces the inquisitive account of uncertainty, comparing it to both inquisitive full belief and to classical uncertainty. In §3.4, I show how inquisitive decision theory helps us theorise about deliberation. In §3.5, I examine the similarities and differences between inquisitive decision theory and the fragmentation accounts of belief developed of Elga and Rayo 2016, 2019 and Yalcin 2018. I conclude that the holistic nature of the present theory clearly distinguishes it from those approaches, and indicate why I think it provides a more compelling solution to the problems of the classical picture.

3.1 Classical Uncertainty

In classical decision theory, an agent's credences form a probability that is defined on every intensional proposition, and this determines an expected value for every possible option:

> A ***classical probability*** is a total function $\mathbf{Pr}$: $\mathscr{P}(\mathscr{W}) \to [0, 1]$ from intensional propositions to the unit interval, subject to the following conditions:
>
> i) *Normalisation*: $\mathbf{Pr}(\mathscr{W}) = 1$
>
> ii) *Additivity*: For any inconsistent (disjoint) intensional propositions p and q,
>
> $$\mathbf{Pr}(\mathrm{p} \cup \mathrm{q}) = \mathbf{Pr}(\mathrm{p}) + \mathbf{Pr}(\mathrm{q}) \tag{3.1}$$
>
> Let $\mathbf{Pr}$ be a classical probability and $\mathbf{a}$ an option. Then $\mathbf{a}$'s ***expected utility*** given $\mathbf{Pr}$ is
>
> $$\mathcal{E}_{\mathbf{Pr}}(\mathbf{a}) := \Sigma_{w \in \mathscr{W}} \mathbf{Pr}(\{w\}) \cdot \mathbf{a}(w) \tag{3.2}$$

As before, I am assuming here that the space of possible worlds $\mathscr{W}$ is finite.

To translate these expected utilities into behavioural predictions, we say that an agent's preferences between options align with the expected utilities. Faced with any decision problem, the agent will be disposed to perform whatever option has the highest expected utility. If there is a tie for highest value, the theory does not say which winner is chosen. If we call that behaviour ***maximising expected utility***, then we can formulate the classical view of credence states as follows:

> **Classical Credences.** An agent α's credences form a classical probability $\mathbf{Cr}_\alpha$, and manifest themselves in a disposition to maximise expected utility with respect to $\mathbf{Cr}_\alpha$ in any decision situation. (3.3)

The classical view of credences (3.3) is closely linked to the classical view of belief states (2.2) and the classical view of individual beliefs (1.10). To see this connection, assume that outright belief or certainty is equivalent to credence 1. I admit that I am cutting some corners in making this identification: it is well known that credence 1 is in fact strictly weaker than absolute certainty, because the two clearly come apart in infinitary contexts (see e.g. Williamson 2007). But the distinction is subtle enough that we can safely ignore it for present purposes, especially since we are assuming anyway that $\mathscr{W}$ is finite.

If we do identify full belief with credence 1, then the classical views of belief states (2.2) and individual beliefs (1.10) are are corollaries of the view about credences (3.3). The definition of a classical probability (3.1) entails that propositions with probability 1 are consistent, and that they are closed under conjunction and entailment. Furthermore, (3.3) associates the same disposition with credence 1 as (1.10) associates with belief:

> $\mathbf{Cr}_\alpha(\mathrm{p}) = 1$ if and only if α is disposed to avoid p-dominated options in any decision problem. (3.4)

For if $\mathbf{Cr}_\alpha(\mathrm{p}) = 1$, $\mathcal{E}_{\mathbf{Cr}_\alpha}(\mathbf{a}) > \mathcal{E}_{\mathbf{Cr}_\alpha}(\mathbf{b})$ whenever $\mathbf{a}(w) > \mathbf{b}(w)$ for all $w \in \mathrm{p}$. Thus (3.3) entails that α will forego p-dominated options if $\mathbf{Cr}_\alpha(\mathrm{p}) = 1$. Conversely, if $\mathbf{Cr}_\alpha(\mathrm{p}) < 1$, (3.3) says α will bet against p whenever the cost is low enough and the potential reward is high enough, and thus will not in general avoid p-dominated options.

That is not to say that (3.3) is *strictly* stronger than (2.2) and (1.10). For interestingly, there is also a sense in which (3.3) is already contained in (1.10). In the last chapter, we saw that the classical view of belief states (2.2) can be viewed a natural consequence of the the classical view of individual beliefs (1.10). The same is true of the classical view of credence states (3.3). The reason for this is that it is implicit in (1.10) that classical believers can never make a sequence of bets that are guaranteed to result in a net loss. Or to use the jargon, they can never be *Dutch booked*.[19] And it is fairly well-known that, modulo some constraints, the only way to immunise yourself from all possible Dutch books is to maximise expected utility with respect to some unique probability (this is theorem 4.12 in Chapter 4 below; see also Ramsey 1926, Lehman 1955, Hájek 2005, Elga 2010). Thus, modulo a few assumptions, the three basic views that constitute the classical picture, the view of individual beliefs (1.10), the view of belief states (2.2) and the view of credences (3.3), are all inter-derivable.

---

[19] Say an option $\mathbf{b}$ is *absolutely dominated* if there is an an alternative $\mathbf{a}$ such that $\mathbf{a}(w) > \mathbf{b}(w)$ for every world $w \in \mathscr{W}$. Now (1.10) says that an agent who believes p is disposed to avoid p-dominated options in every choice they make. Since absolutely dominated options are p-dominated, it follows that any classical believer will avoid absolutely dominated options in every choice they make. In Chapter 2, I clarified that this is supposed to cover composite as well as simple choices. Thus (1.10) entails that no classical believer will make a sequence of choices that is guaranteed to lose them utility.

3.2 Logical Omniscience and the Preface Paradox

Even though the classical view of credence states (3.3) in some sense entails the classical view of outright belief states (2.2), it is not obvious that every problem I raised for (2.2) should also affect (3.3). For (3.3) matches up with (2.2) in that classical *certainties* are consistent and closed under conjunction. But now consider the set of all intensional propositions in which an agent is *confident*, in the sense of having, say, credence 0.8 or higher. This set will ordinarily be inconsistent and it will not be closed under conjunction. So there is at least a prima facie appearance that the shift from full beliefs to credences could account for some of the phenomena we have been concerned with in this dissertation.

The preface paradox makes this concern vivid (Makinson 1965). A history professor is confident about, and willing to act on, every single one of the historical claims $p_1, p_2, \ldots, p_{853}$ she makes in her lengthy book. But in the book's preface, she expresses confidence that not all the claims are true, which is to say that $\neg\bigcap_i p_i$. If all you have are classical information states, you cannot straightforwardly account for the professor's attitudes. But once credences are on the table, the problem goes away. Classical probabilities can attach a high probability to each conjunct p_i while also assigning a low probability to their conjunction $\bigcap_i p_i$. However, the move to credences does not address the sorts of issues about memory retrieval and deductive failure that have occupied us in this dissertation. To see why, I need to say more about the way of the classical idealisations resurface in context of the classical view of credences (3.3).

For a start, the move to credences certainly will not help account for the hyperintensional contrast that the original ROMEO RECALL and TRIVIAL TROUBLE cases highlighted. After all, the objects of classical credences are still intensional propositions, and in these cases only a single piece of information is in play: respectively, *Romeo's number is 212-529 6300* and *"dreamt" ends in -MT*. In each example, the agent acts on this intensional proposition in one situation, and fails to act on it in another. No matter what credence we say they have in this proposition, (3.3) is not going to account for that contrast. At most, (3.3) suggests that Tom of TRIVIAL TROUBLE should be *more* prone to give the sphinx the right answer than to spell his letter correctly, since the stakes are higher in the former situation.

Closure under single-premise entailment in (2.2) is mirrored in (3.3) by the fact that any credence state **Cr** assigns greater probability to weaker propositions: whenever p entails q, **Cr**(p) ≤ **Cr**(q). So given (3.3) it is not just impossible to be *certain* that Detroit had 200 murders without also being certain that Michigan does, but also impossible to deem it *likely* Detroit has a high murder rate without deeming it *at least as likely* that Michigan does (as before, I am treating this as a single-premise entailment, ignoring the possibility that Detroit is in a different state). Mandy from MITTEN STATE MURDERS I acts only on the former proposition. But (3.3) wrongly predicts she should be equally or more prone to act on the latter proposition. Relatedly, (3.3) cannot adequately capture Mandy's mental state in MITTEN STATE MURDERS II either: when p and q are inconsistent, one cannot have high credence in both, since **Cr**(p) ≤ **Cr**(¬q).

Multi-premise entailments are where things get more interesting. Again, (3.1) links the likelihood of the conclusion to that of the premises, but this time the connection is weaker:

If $p_1, p_2, \ldots, p_n$ entail c, then $\mathbf{Pr}(c) \geq 1 - \Sigma_i \mathbf{Pr}(\neg p_i)$.[20] (3.5)

Crucially, (3.5) does *not* guarantee the transmission of confidence in the premises p_i to confidence in the conclusion c. Confidence in p_i guarantees that $\mathbf{Cr}(\neg p_i)$ is small, but those small numbers can add up. However high confidence in the premises is, as long as there are a lot of them, it is compatible with (3.5) that the credence in c is low, even when that means that the agent is confident in an inconsistent set of propositions $\{p_1, p_2, \ldots, p_n, \neg c\}$. That is why the classical view of credence (3.3) has an easy treatment for the preface paradox, which turns essentially on the multitude of propositions involved.

So the interesting question is how much sense classical decision theory can make of an agent's failure to see the conclusion of a many-premise entailment, as in the following example:

> SILENT SUDOKA: Expert sudoku solver Sarah is shown The World's Hardest Sudoku (Hutchinson 2010). After studying the puzzle, with its twenty-one clue entries, for some amount of time, she is offered a prize if she can say in which column the "4" in the fourth row goes. Not having made much progress with the puzzle, Sarah guesses the "4" goes in column 7; but in fact the correct answer is column 3.

On the face of it, SILENT SUDOKA does not conflict with (3.5). Twenty-two pieces of information are needed to derive the correct solution, if we count each clue entry as separate puzzle pieces and the rules of sudoku as the final piece. Call those propositions the *premises* $e_1, e_2, \ldots e_{22}$. These premises entail c_3, where c_n is the proposition that *the "4" in the fourth row goes in the n^{th} column*.

[20] *Proof*. Note $\mathbf{Pr}(q \cap p) \geq \mathbf{Pr}(q) - \mathbf{Pr}(\neg p)$; by induction we get $\mathbf{Pr}(\bigcap_n p_n) \geq 1 - \Sigma_n \mathbf{Pr}(\neg p_n)$; the inequality (3.5) then follows from the single premise closure property: $\mathbf{Pr}(c) \geq \mathbf{Pr}(\bigcap_n p_n) \geq 1 - \Sigma_n \mathbf{Pr}(\neg p_n)$.

Even if Sarah memorised all the clues, and is 95% confident in each e_i, then for all (3.5) says, her credence in c_3 could be 10%, or even 0. So (3.3) does have a way to account for Sarah's choice.

On closer examination, there are in fact two possible classical explanations. But they are not what you might expect. The obvious explanation for Sarah's inability to identify the right column is related to the fact she has not solved the sudoku. In particular, she has not worked out any connection between the twenty-two pieces of information she was given, and the position of the '4' in the fourth row. After all, this is an extremely difficult sudoku that would take even an expert solver days to solve. However, this kind of explanation is not available to the classical theory, which needs to work around the fact that Sarah's conditional credence in c_3 given the premises $\bigcap_i e_i$ is inevitably equal to 1.

That leaves two possibilities. One is that, for whatever reason, Sarah's credences in the premises e_i are not independent, and that she perceives a strong tension between them. For instance, Sarah could be *certain* that their conjunction $\bigcap_i e_i$ is false, but be unsure which premise e_i is false. On this explanation, Sarah's confidence that $\bigcap_i e_i$ is false means she puts no stock in its consequences either. The other possibility, which overlaps with the first, is that Sarah has a strong prior leaning towards column 7 specifically, so that her credence in c_7 outweighs her above-average credence in c_3.[21]

---

[21] (3.3) predicts Sarah will choose column 3 over column 7 if $\mathbf{Cr}(c_3) > \mathbf{Cr}(c_7)$, where $\mathbf{Cr}$ is her credence function. If Sarah perceives no tension between the premises, her credences in the premises are at least independent. Thus $\mathbf{Cr}(c_3) \geq \mathbf{Cr}(\bigcap_n e_n) \geq 0.95^{22} \approx 0.32$, because $\mathbf{Cr}(c_3 \mid \bigcap_n e_n) = 1$. But then $\mathbf{Cr}(c_3) > \mathbf{Cr}(c_7)$ unless $\mathbf{Cr}(c_7 \mid \neg\bigcap_n e_n) > 0.48$. So c_7 would have to be over eight times more likely than the other eight columns, conditional on $\neg\bigcap_n e_n$.

Now these are both possible explanations of Sarah's choice of column 7. Perhaps she has special reason to think she is mistaken about one of the clue entries, without knowing which one. Or maybe somebody lied to her, telling her c_7 was the answer. The case as stated does not specify. But assuming Sarah has in fact not solved the sudoku, neither of those explanations are required. If the Sarah were a classical agent, her choice of column 7 would reveal that she must either be confident in $\neg\bigcap_i e_i$ or in c_7, and must therefore be disposed to accept low betting odds on one or both of these propositions. But in actual fact, this need not be the case. So classical decision theory can only account for Sarah's choice of column by distorting other aspects of her behavioural dispositions. This matches the general pattern we have seen in all our case studies: the classical story may have a way to account for each individual choice, but fails to capture the overall behavioural pattern.

Still more tellingly, the classical treatment of credences entails that Sarah's uncertainty about c_3 is essentially connected to her uncertainty about the clue entries. If that were true, Sarah could increase her confidence in c_3 simply by refreshing her memory of the clue entries e_i. Thus the classical view predicts that finding the solution to a sudoku is a matter of memorising its clues by heart. But anyone who has ever got stuck on a logic puzzle knows that in fact this is not how it works.

It is different in the paradoxical preface case. The professor's confidence that her book contains a mistake *is* directly connected to her uncertainty about its individual claims. At least in principle, addressing her doubts about each and every one should eventually lead her to doubt whether there really is a mistake. Conversely, if she finds evidence showing p_{639} to be false, then that discovery will bolster her confidence in $\neg\bigcap_i p_i$. So her beliefs about the conjunction and the conjuncts are connected in a way that Sarah's credences about c_3 and the entries e_i are not. Because the professor's attitudes are linked in this way, they can still plausibly be viewed as aspects of a single, overall world view, as the classical picture demands. Sometimes it is claimed that the preface paradox and related examples show that the classical picture allows for doxastic inconsistencies and failures of deductive closure (Kyburg 1970, Hawthorne 2009; cf. Leitgeb 2014). That might be true in some sense, but this is not the sort of deductive closure violation that interest us here.

By contrast, Sarah's credences *are* disconnected. The classical account (3.1-3) demands that Sarah's confidence about the clue entries e_i and her credences in c_3 should linked in a way that respects the entailment relation that holds between them. But Sarah's credences are not linked in this way, because she does not see the logical connection between these propositions. Even in its probabilistic incarnation, the classical view rules out this kind of cognitive disconnect. So SILENT SUDOKA, on the most natural interpretation, demonstrates a very different kind of closure failure than the preface paradox. Deductive failures of that kind are every bit as much at odds with the classical picture as the other phenomena adduced in this dissertation.

3.3 Inquisitive Uncertainty

Now for the inquisitive account of partial belief. Like an inquisitive belief state, an inquisitive probability has a domain of questions; it assigns probability values to every quizposition about a question in its domain:

Let $\mathscr{D}_{\mathbf{Pr}}$ be a non-empty set of partition questions closed under parthood. Then an ***inquisitive probability*** on $\mathscr{D}_{\mathbf{Pr}}$ is a function $\mathbf{Pr}$: $\mathscr{Q}(\mathscr{D}_{\mathbf{Pr}}) \to [0, 1]$ from quizpositions about questions in $\mathscr{D}_{\mathbf{Pr}}$ to the unit interval, such that for any Q, R $\in \mathscr{D}_{\mathbf{Pr}}$:

i) *Normalisation*: $\mathbf{Pr}(Q^Q) = 1$

ii) *Additivity*: If Q contains R, A^Q and B^R are inconsistent, and C = A ∪ QB, then

$$\mathbf{Pr}(C^Q) \;=\; \mathbf{Pr}(A^Q) + \mathbf{Pr}(B^R). \tag{3.6}$$

Recall that by definition (2.7) above, $QB =_{df} \{ (q \cap b) : q \in Q \text{ and } b \in B\} \setminus \{\varnothing\}$. The notation "$\mathscr{Q}(\mathscr{D}_{\mathbf{Pr}})$" represents the space of all quizpositions A^Q such that $Q \in \mathscr{D}_{\mathbf{Pr}}$.[22] Note that, within any given question $Q \in \mathscr{D}_{\mathbf{Pr}}$, the probabilities $\mathbf{Pr}(A^Q)$ are just like a classical probability, because we have both normalisation $\mathbf{Pr}(Q^Q) = 1$, and additivity: $\mathbf{Pr}(A \cup B)^Q \;=\; \mathbf{Pr}(A^Q) + \mathbf{Pr}(B^Q)$ whenever A, B ⊆ Q are disjoint. Call this subprobability **Pr**'s ***probabilistic view*** about Q. Whenever R is part of $Q \in \mathscr{D}_{\mathbf{Pr}}$, **Pr**'s probabilisitic view on Q settles its probabilisitic view on R: the probability of every quizposition about R is equal to the probability of the intensionally equivalent quizposition about Q. To see this, note that if A^Q and B^R are intensionally equivalent, then by (ii), $\mathbf{Pr}(A^Q) = \mathbf{Pr}(\bot^Q) + \mathbf{Pr}(B^R) = \mathbf{Pr}(B^R)$, since $\mathbf{Pr}(\bot^Q) = 0$.

It follows that probabilistic views on overlapping questions must agree on the overlapping part. Just like outright, full beliefs, the partial beliefs of an inquisitive agent are constitutively linked through their thematic connections, forming a web of credences. In general, the more questions its domain $\mathscr{D}_{\mathbf{Pr}}$ contains, the better integrated **Pr**'s probabilistic views are. And at the limit,

---

[22] Geeky side-note: the spaces $\mathscr{Q}(\mathscr{D})$ of quizpositions on a domain $\mathcal{D}$ are technically identical to the incoherent information state on the corresponding domain $\mathcal{D}$. Likewise, classical probabilities are defined on $\mathscr{P}(\mathscr{W})$, which is the inconsistent classical information state.

where $\mathscr{D}_{\mathbf{Pr}}$ is the space of all partition questions, **Pr** reduces to a classical probability.

Adapting the classical definition in the natural way, we get the following inquisitive notion of expected value:

> Let **Pr** be an inquisitive probability, let Δ be a choice raising a question $Q \in \mathscr{D}_{\mathbf{Pr}}$, and let **a** be an option in Δ. Then **a**'s ***expected value*** given **Pr** is
>
> $$\mathcal{E}_{\mathbf{Pr}}(\mathbf{a}) \;:=\; \Sigma_{q \in Q}\, \mathbf{Pr}(\{q\}^{Q}) \cdot \mathbf{a}(q) \tag{3.7}$$

To check that this is well-defined, we need to establish that when two questions Q and R address the same choice Δ, they both yield the same expected value for every option in Δ. To see this, recall from §2.11 that whenever Q and R address Δ, they always share a part S that addresses Δ. Using inquisitive additivity, we can then see that for any option $\mathbf{a} \in \Delta$,

$$\Sigma_{q \in Q}\, \mathbf{Pr}(\{q\}^{Q}) \cdot \mathbf{a}(q) \;=\; \Sigma_{s \in S}\, \mathbf{Pr}(\{s\}^{S}) \cdot \mathbf{a}(s) \;=\; \Sigma_{r \in R}\, \mathbf{Pr}(\{r\}^{R}) \cdot \mathbf{a}(r).$$

Thus an inquisitive probability **Pr** yields an unambiguous assignment of expected values to the options of any decision problem that raises a question in $\mathscr{D}_{\mathbf{Pr}}$. And that allows us to state the inquisitive view of credences in the following way:

> **Inquisitive Credences.** An agent's credences form an inquisitive probability **Cr**, manifesting themselves in a disposition to perform actions that maximise expected utility with respect to **Cr** in any decision situation that raises a question in $\mathscr{D}_{\mathbf{Cr}}$. (3.8)

Just as in the classical case, we can show that (3.8) entails the inquisitive views of belief states (2.12) and of individual beliefs (1.12), provided that credence 1 is interpreted as an outright belief.[23] There is an analogue in the converse direction as well. Given suitable assumptions, and given an interpretation of the inquisitive view of individual beliefs (1.12) according to which it covers composite choices, it can be shown that any agent with at least one inquisitive belief in the sense of (1.12) must maximise expected utility with respect to some inquisitive utility. One of the main aims of Chapter 4 is to prove this claim, theorem 4.18.

Since the inquisitive view of credences (3.8) incorporates the inquisitive view of belief states (2.12), it is clear in advance that (3.8) also has the resources to explain what is going on in the various problem cases we discussed. But to get a better sense of how the theory works, it may be helpful run through them all the same, beginning with the MITTEN STATE MURDER cases.

It follows directly from (3.6)'s *Additivity* that parts are always more likely than the whole. This is the probabilistic analogue of closure under parthood. But it is compatible with (3.6) that some entailments should be less likely than the entailing proposition — that is the probabilistic analogue to failure of closure under entailment. So since the questions M, *the number of Michigan murders* and D, *the number of Detroit murders*, are disjoint, Mandy's credences about those two questions are in principle independent. In particular, Mandy could have high credence in the

---

[23] *Proof*. For the reasons explained in §3.1 above, (3.8) associates the same disposition with credence 1 as (1.12) does with full belief. Now let **Cr** be any inquisitive credence and let

$$\mathbf{B} := \{ A^Q : \mathbf{Cr}(A^Q) = 1 \}.$$

Since $\mathbf{Cr}(\bot^Q) = 0 \neq 1$, **B** is *coherent*. It follows from *Additivity* that $\mathbf{Cr}(A^Q) = \mathbf{Cr}(A\neg B^Q) + \mathbf{Cr}(B^R) \geq \mathbf{Cr}(B^R)$ when B^R is part of A^Q, whence **B** is *closed under parthood*. Now suppose $\mathbf{Cr}(A^Q) = \mathbf{Cr}(B^R) = 1$ and R is part of Q. Then $\mathbf{Cr}(A\neg B^Q) \leq \mathbf{Cr}(Q\neg B^Q) = \mathbf{Cr}(\neg B^R) = 0$. So by *Additivity*, $\mathbf{Cr}(AB^Q) = \mathbf{Cr}(A^Q) - \mathbf{Cr}(A\neg B^Q) = 1 - 0 = 1$. Thus **B** also satisfies *partial closure under conjunction*, making **B** an inquisitive belief state. Hence, if we identify credence 1 with belief, (3.8) entails both (2.12) and (1.12).

proposition *Detroit has* ***more than 150*** *murders* while simultaneously having arbitrarily high credence that *Michigan has* ***less than 150*** *murders*. Furthermore, she can adjust either view without affecting the other at all. On the classical theory, the entailments involved produce a necessary connection between Mandy's credences about these two topics. There are no such necessary connections given the inquisitive theory, unless her credences about these two questions happen to be integrated in some larger view.

The same basic features of the inquisitive account allow us to handle the SILENT SUDOKA. There are a hundred different ways in which Sarah might memorise the clue entries. One natural way to do it is to memorise, for each 3×3 block, the answers to *What pattern do the clues in this block form?* and *In order, what numbers are clued in this block?* If Sarah makes sure she has high credence in the complete answers to those 18 questions, it will allow her to reproduce all the clues. But since none of these questions overlap with the question *In which column does the 4th row "4" go*, doing so will not itself affect her credences about that issue. This way, the inquisitive model can explain why the best Sarah can do is to venture a random guess about the column, in spite of the fact that she has all the necessary information to infer the correct answer.

Let us now ask a different question: how might Sarah go about finding out in which column the "4" goes? We saw that the classical picture suggested a bizarre approach: simply memorise all the clue entries until you are almost certain of each one. Once you are done with that, the answer should simply jump out at you. Does the inquisitive picture suggest a better strategy? In theory, one possibility would be to bring all the clues to bear on the question *What is the distribution of numbers in the grid*. If Sarah managed to accomplish this, she would basically be like the classical agent, and see the solution instantly. However, that strategy is not feasible in this case. As discussed in §2.7, one pays a significant cognitive price for maintaining views on large questions, which require a high level of integration. This particular question has $2 \cdot 10^{77}$ cells. Obtaining a view on it would require cosmic levels of computational power, and is a merely theoretical possibility.

Thankfully, the inquisitive picture also suggests some strategies that do not require superhuman levels of intelligence. Sarah can try to construct inquisitive daisy-chains of views of the kind discussed in §2.6-10, with the aim of linking the views she started with to her target question. Doing so will involve forming views on a long series of strategically posed questions. As she moves down the chain, she may forget some views after they have served their purpose, preserving only the conclusions. During this deductive process, Sarah's views about different corners of the puzzle continuously evolve, standing in shifting mereological relations to one another. It need never happen that Sarah oversees the whole long chain of inferences that leads from the clue entries to the controversial "4".

This is just a bare sketch of how the sudoku solving might be modelled on the inquisitive picture; a more detailed treatment would go into specifics about the information states traversed and the questions posed. But I hope I have said enough to convey the basic idea, and I hope that the emergent picture of the deductive process looks recognisable. It certainly seems to resemble real-world puzzle solving strategies a great deal better than the bizarre memorisation strategy suggested by the classical picture.

Finally, I want to have another look at the old recognition/recall cases, which bring out an important contrast between inquisitive and classical decision theory. The basic thought is the same as it always was. In ROMEO RECALL, Juliet gives high credence to the quizposition ***Romeo's*** *number ends in -6300*, but not to *Romeo's number ends in* ***-6300***. In TRIVIAL TROUBLE, Tom gives high credence to the quizposition *"dreamt" ends in* ***-MT***, but not to the quizposition ***"dreamt"*** *ends in -MT*. But just what are Juliet and Tom's attitudes to the second quizposition? It is not like they are certain it is false, either, or that they reckon it is an even bet.

In fact, (3.8) dictates a particular answer. Given that we do not want to say that Tom has *high* credence that ***"dreamt"*** *ends in -MT*, we must say he lacks a credence about that quizposition altogether. The reason is that, whenever an inquisitive probability is defined over two intensionally equivalent quizpositions, it always assigns them equal probability. For suppose some questions $Q, R \in \mathscr{D}_{\mathbf{Pr}}$ have intensionally equivalent answers A^Q and B^R. Say $p = \bigcup A = \bigcup B$. Then Q and R overlap at least on the polar question $P = \{ p, \neg p \}$. It follows that

$$\mathbf{Pr}(A^Q) \quad = \quad \mathbf{Pr}(B^R) \quad = \quad \mathbf{Pr}(\{p\}^P).$$

So on the inquisitive account of his behaviour, Tom has *no* view on the question *Which English words end in -MT*; likewise, Juliet has *no* view on the question *What are the last four digits of Romeo's number*. This is supposed to explain why these agents, when confronted with those questions, are unable to respond appropriately. Intuitively, the absence of a view to guide their decision is a pretty satisfactory explanation for their behaviour.

However, from a more theoretical standpoint, something is missing here. With the exception of ties, classical decision theory makes a prediction about the agent's response to every decision situation: a classical probability assigns expected values to all the options of every choice. But an inquisitive probability only returns expected values when the choice raises a question in its domain. So formally speaking, inquisitive decision theory, in the form of (3.8), makes no predictions at all about agents' choices when facing a question on which they have no view. In particular, it does not make a positive prediction about what Tom and Juliet will do in these situations.

3.4 Deliberation

Thus inquisitive decision theory has a problem that does not arise for the classical theory: how do agents respond to new questions, on which they do not have any view at all? For all (3.8) says, Tom's response to the decision problem Δ_{Sphinx} on p. 29 could be anything at all, given that he has no view on the question the sphinx asks him. This suggests the replies "carrot" and "elephant" would be equally in accord with Tom's mental state as the reply "I don't know". But clearly, that is not right. There is an obvious, belief-based explanation for why Tom did not reply "elephant": Tom knows "elephant" ends in -NT, not on -MT. And he does not want to embarrass himself by suggesting otherwise. In this section, I show that on a more careful analysis, it turns out that inquisitive decision theory can in fact underwrite this explanation. This is thanks to the fact that the thought process by which Tom arrived at this choice can itself be viewed as consisting of a series of simpler choices.

On the inquisitive picture, a course of action can sometimes be described as the product of rational and belief-guided decision-making, but only when it is viewed as a sequence of choices rather than a single choice. One example of such a course of action came up several times in this chapter already: filling in a sudoku grid with the aim of arriving at the correct solution. In doing so, you in effect select one of a gazillion possible options of putting numbers in the grid. This is a hugely complex decision-problem, raising the vast question *What is the distribution of numbers in the grid*. But your response to that decision problem is not based on an antecedent view about that question: for the reasons discussed, it is not humanly possible to have such a view. Nonetheless, we can understand your course of action as belief-guided and rational if we split it up into parts. Rather than being the product of a single atomic choice, the way you filled in the sudoku is the product of a complex deliberative process that is at each stage guided by your beliefs on smaller questions. Analogous points can be made about Rubik's cube example in §2.7.

In general, complex choices tend to be composed of smaller, simpler ones. Walking to work, you do not need to decide the entire route when leaving home: you can take it one turn at a time. Writing a poem, you need not decide on the entire text before putting your pen to the paper: you can write it one line at a time. Or take the biggest decision problem of all, *How to live?* Classical decision theory suggests you could just pick a hyperplan in infancy and stick with it for the rest of your life. The inquisitive picture explains why this is not what we do: the question involved is far too complex for our finite minds. We must instead live life one choice at a time. The smaller choices of which larger choices are composed typically raise smaller questions about which one can more easily have beliefs. If inquisitive decision theory succeeds in correctly predicting the agent's move at every turn, I contend it succeeds in explaining the behaviour successfully: no need to cut corners by trying to do it in one go.

Let's first see how this perspective helps with ROMEO RECALL. I pointed out that in the phone booth, having dialled the first six digits of Romeo's number, Juliet faces a question on which she has no view: *What are the last four digits*. But she is also, simultaneously, in a different, simpler decision problem, which raises a simpler question: namely the problem of dialling the next digit. Perhaps she does have a view on the question that problem raises, *what is the seventh digit of Romeo's number*. Supposing the view divides her credences equally between the ten possibilities, we can simply explain her action (leaving the phone booth, say), as being based on her uncertainty about the seventh digit.

What if Juliet does not have an antecedent view on *what the seventh digit of Romeo's number is* either? Well, a view on this smallish question can at least be acquired at relatively moderate cognitive cost. So in that case we might say that Juliet first did some thinking, then concluded she was unsure about the next digit, then decided to leave the phone booth. Her decision to do some thinking can itself also be understood as a response to a decision situation, in which she had a choice between giving up right away and thinking it over. On the classical picture, there would be no point in thinking it over: if Juliet had the relevant information, she would have it immediately available. But the inquisitive picture makes sense of the possibility she might have the information, without its being immediately present to mind. Thus it makes sense of Juliet's reasonable expectation that racking her brain might yield the answer she seeks. We can thus account for Juliet's metacognitive decision to do so as a regular inquisitive belief-guided choice. Thus we can recast what appeared like a response to a question on which Juliet has no view as a series of responses to simpler questions, each one of which is guided by Juliet's beliefs at the time of the choice.

Juliet's decision to think about the question she faces before responding is an example of a kind of default choice we make very frequently. Compare it to the default expectation in conversation that interlocutors ought to think about the questions they are asked before replying. Suppose someone asks you *What day of the week Judy's birthday is*. You have no view on the matter and have not considered the issue before. Rather than issuing a reply at random, you would ordinarily think about it for a second first. Having thought about it, you may find the answer to the question, or you may be unsure: either way you have now acquired a view on the matter. The explanation of your reply is based on the view you acquired after cogitation. But really the deliberative process involved making two choices: first a choice to think about the question, and then a choice of reply.

So the proposal is to view Juliet's deliberative process as consisting of a series of decisions that is guided in part by her metacognitive beliefs. In psychology, this metacognitive view of deliberation is popular, and there is a lot of evidence for it: see e.g. Proust 2013, chs. 1, 2, 5. The energy and time one spends on a mental search like Juliet's depends on the stakes, and on her confidence that she will arrive at an answer. It turns out human beings are good at making such estimates. Research on metamemory has shown that human subjects frequently have a certain "feeling of knowing" (FOK), and that this sensation is is a remarkably reliable indicator of the information that is in fact stored in memory (Hart 1965, Murray and Tiede 2008).

A proper treatment of Tom's reply to the sphinx in TRIVIAL TROUBLE requires something a little more sophisticated, since the mental process was probably more interesting. Instead of conceiving of Tom's options as his ultimate replies to the sphinx, conceive of these replies as the potential end points of a mental quest to find an answer to the sphinx's riddle. The aim is to explain his choices at each turn in the quest. At the start, Tom can decide to either *search* his memory for words end on -MT, or to *give up* right away and say "I don't know". What he does there depends on his confidence that the search will succeed. Unless Tom has very little faith in himself, he will at least give it a try. Let us say the first word that comes to mind is the word "tempt". He now faces something like the following choice:

	/tɛmt/ ends in -MT	/tɛmt/ ends in -MPT, but search will succeed	/tɛmt/ ends in -MPT, and search will fail
reply /tɛmt/	1	–1	–1
continue search	$1 - C$	$1 - C$	$-C$
give up	0	0	0

TABLE 7: TEMPTED?

Here C is the cost associated with continuing the search. Tom may not antecedently have a view on the three-partite question that this choice raises, but it is an easy deduction away. After all he does know how "tempt" is spelt, and so he has an answer to the binary question *Does "tempt" end on -MT* — all he has to do is combine this with his metacognitive view on the chances of a successful search, and he is ready to make this choice. If he continues the search, he will after some time come up with another candidate, like "exempt". At that point, he faces a similar decision problem again. Unless Tom hits upon the word "dreamt", his credence that the search will be successful should go down with each cycle of the process. And if the sphinx is getting impatient, the cost C of a search goes up. Eventually, the expected value of continuing the search will turn negative, at which point (3.8) predicts Tom will give up.

At every juncture of this process, Tom is responding to a question to which he has an answer at that time. His beliefs about spelling feed into the process, ensuring he will not give an answer that he knows to be wrong. In particular, this guarantees that he will not say "elephant" or "carrot". Assuming his beliefs about spelling are true, there are in fact only two actions that can come out of this deliberative process given Tom's starting point: either he answers "dreamt" or "I don't know". Which of those two replies he ends up giving remains indeterminate, since we are treating the outputs of each word search as more or less random. But that is as it should be: it is in fact unpredictable whether Tom will get lucky and stumble on the right word.

In redescribing Tom's response to Δ_{Sphinx} as the product of a series of smaller choices, I do *not* mean to suggest that the expected response to a decision problem can vary depending on how the decision is described. That would undermine the predictiveness of inquisitive decision theory. Rather, the possibility of dividing the deliberative process into smaller parts should make the theory *more* predictive, because it allows us extract predictions about cases on which the theory at first glance appears to be silent. Tom still counts as having made decision in Δ_{Sphinx}. But since he did not have any views that addressed this decision problem, inquisitive decision theory does not directly yield a prediction about that choice. Even in cases like that, the agent's response can still be made intelligible using inquisitive decision theory, by appreciating that, *in addition* to being a response to Δ_{Sphinx}, Tom's reply was also the product of a sequence of smaller deliberative choices.

3.5 The Atlas or the Web?

Inquisitive decision theory incorporates insights from Adam Elga and Agustín Rayo's fragmented decision theory and also from Seth Yalcin's account of belief as a question-sensitive attitude (Elga and Rayo 2016, 2019, Yalcin 2018; see also Yalcin 2008, ch. 3, 2011, §5-8, and Yalcin 2007, §7 on inquisitive credences). These are both fragmentation theories of belief, building on suggestions from Stalnaker 1984, 1999a and Lewis 1982. Other notable endorsements of views in the fragmentation tradition include Fagin and Halpern 1988, §6, Braddon-Mitchell and Jackson 2007, Egan 2008, Greco 2015b, Kindermann, Borgoni and Onofri fc. In this section, I examine the similarities and differences between those fragmentation theories and the inquisitive theory of belief and action articulated over the past three chapters. In particular, I explain why I think the latter is *not* a fragmentation theory, except in the superficial sense that it invalidates general closure under conjunction. For simplicity, I will focus here on the comparison between inquisitive (full) belief states and fragmented belief states. But I take it that everything I say applies *mutatis mutandis* to fragmented credence states as well.

In Seth Yalcin's helpful analogy, fragmentation theories hold that beliefs do not form a map but rather an *atlas*: a collection of maps. Instead of always steering by the same map, we change it up from time to time, steering sometimes by one map, and sometimes by another. Each map in the atlas represents a coherent view. But since the contents of distinct maps are in principle independent of one another, one map may contain blanks in regions where another map is richly detailed. The different maps can also contradict each other.

On Stalnaker's and Elga and Rayo's version of this view, every map in the atlas is a map of the entire world. That is, according to their theory, a single agent is in multiple classical belief states or classical credence states at once. These belief states are separate from one another and can in principle vary independently: Stalnaker and Lewis call this *compartmentalisation*. In particular, that means that the various belief states of an agent can be inconsistent with each other. The agent counts as believing p just in case p is included in one of their belief states. Thus the theory invalidates closure under conjunction: an agent may believe p, and also q, but if p and q happen to be in separate fragments, they need not believe $p \cap q$. But this version of the view does not invalidate closure under entailment: if an agent believes p, they believe everything p entails.

On Yalcin's version of the view, each map in the atlas represents a different region of the world. Formally, each belief state is associated with a distinct partition question Q, and its contents are some Q-answer V.[24] The entire atlas can be represented as a partial function from questions to answers. Yalcin's belief states are similar to the inquisitive views from §2.5, in that the only entailments of V that the agent believes are the Q-answers A such that $V \subseteq A \subseteq Q$. Thus Yalcin's fragmentation theory invalidates not only general closure under conjunction but also general closure under entailment. In addition, Yalcin's proposal invalidates the analogues of closure under parthood and partial closure under conjunction, since he imposes no constraints on the relationship between believers' views on different questions.

The similarities between these fragmentation accounts and inquisitive decision theory are pretty

---

[24] I am omitting the superscript Qs here because in Yalcin's theory, the objects of belief are still intensional propositions, not quizpositions: the question answered is not part of the content of a belief.

obvious. Inquisitive beliefs can contradict each other, but they cannot be contradictions. Inquisitive belief states invalidate conjunctive closure in cases where the conjuncts belong to distinct views, and invalidate closure under entailment where an agent's information is not carried over from one view to the next. So the plurality of inquisitive views is playing a somewhat similar role to the plurality of belief states in fragmentation theories. At the same time, there are clear differences, even at the abstract, logical level. Unlike belief fragments, inquisitive views are not at all separate. There is no compartmentalisation. Inquisitive views overlap with one another and are constitutively tied together through their thematic connections. That is why inquisitive belief states are much more akin to a web than to an atlas with cleanly separated pages. According to fragmentation theories, an agent can in principle believe absolutely any combination of contingent propositions. On the inquisitive theory, for instance, agents cannot believe contradictories (that is, they cannot have two beliefs equivalent to one another's negation).

There is a deeper reason for those differences. It is part of the core idea of fragmentation that one only uses only one map at a time. So a belief can be "on" or "off" depending on whether it is included in the fragment that happens to be in charge. But on the inquisitive picture, all our beliefs are always on. It is never the case that one view dictates our action at the exclusion of our other views, so that every choice we make is in principle guided by our doxastic state in its entirety.

Sure, a belief A^Q only manifests itself in choices that Q addresses. But that is not fragmentation. That just means that the belief only guides those choices on which it has a bearing, and where it is relevant. Even on the classical view, my belief that Quito is the capital of Ecuador and my belief that Napoleon was short do not typically manifest themselves while I am mowing the lawn or choosing an ice cream flavour. Everyone agrees that beliefs only make a difference in situations where they are relevant. But the fragmentation story has it that a belief should have an effect on action only when it is both relevant and also "active," that is part of the operative belief fragment. In opposition to this, the classical and inquisitive theories agree that our beliefs guide our choices *whenever* they are relevant. It is just that the inquisitive account employs a more sophisticated, hyperintensional notion of relevance.

This matters a great deal, because it goes to the heart of what we have been trying to preserve from the classical picture: namely its capacity to account for the robust and unifying role that beliefs play in explaining our actions. To bring this out, consider an example:

> WILL OF THE WHIST. Will is a novice whist player. His five remaining cards are lying face down on the table in front of him (he can look at them whenever he likes; Will keeps them face down only to hide them from the other players). Will knows for sure that the top card is a trump card, but he does not quite remember what his other cards are. He really wants to play his trump card. But if he has any diamonds, he is obliged to follow suit instead. What will Will do?

Of course, what Will is going to do depends on his beliefs. If he is certain he has no diamonds, he will play his trump card. If he is certain he has diamonds, he will follow suit. If he is unsure, he will probably have a peek at his cards before playing.

But what will he do if those things are all true at the same time, as the fragmentation picture allows? Let us examine this question from the perspective of the Elga/Rayo theory first. Suppose Will's mental state comprises at least these four classical belief fragments:

- A belief state $\mathbf{B}_F$, which entails that Will has a three of diamonds and must follow suit.
- A belief state $\mathbf{B}_T$, saying that Will has no diamonds and is free to play the trump card.
- A belief state $\mathbf{B}_L$, which says Will has only one red card, but is neutral as to whether it is a five of diamonds or a five of hearts.
- A belief state $\mathbf{B}_G$, which is entirely comprised of third-personal geographical facts. It contains a lot of information about African countries and capitals, but nothing about the card game at all. It does not even include the proposition *Will is playing a card game*.

Classically, you would predict on the basis of $\mathbf{B}_F$ that Will would *Follow suit*, on the basis of $\mathbf{B}_T$ that he will *play the Trump card* and on the basis of $\mathbf{B}_L$ that he will *Look at his cards*. It is not clear what the classical theory might predict on the basis of an exotic belief state like $\mathbf{B}_G$. But if $\mathbf{B}_G$ really were his belief state, I imagine Will's priorities would include figuring out where he is, why he is holding five pieces of cardboard and who all these people staring at him even are.

What can we predict about Will's behaviour given that he is in this fragmented doxastic state? Well, nothing, really. Clearly he cannot be having all of these reactions at the same time. The best we can do is to say that Will's action depends entirely on which belief state happens to be active. But absent an account of the conditions under which a belief state may be expected to become activated, that really is no help at all. Here is another way to think of it: the fact that the fragmentation account has an "explanation" ready for every action Will could possibly take, shows that those purported explanations are at best incomplete.

Elga and Rayo offer a rather brute way to fill this explanatory gap. On their version of the account, each belief fragment is accompanied by an *elicitation condition* that specifies the circumstances under which it becomes active. Crucially, the elicitation conditions associated with distinct fragments are always mutually exclusive; thus they partition (some portion of) the space of decision situations. A list of the agent's belief states does not yield behavioural predictions or explanations by itself. According to Elga and Rayo, the true explanans of an agent's behaviour is their *access table*, which is a full specification of all the agent's beliefs fragments together with their elicitation conditions.

So in Will's case, we first need to work out which (if any) of Will's belief states is *elicited* in the particular situation described. Only then will we have a prediction about his behaviour. Depending on the elicitation conditions associated with those four belief states, any of the reactions described is in principle possible. One complaint about this style of explanation is that the elicitation conditions do not really explain much. Rather than saying *why* a certain belief state was "elicited" on a particular occasion, the access table simply lists the occasions when the state is activated. If we are happy with explanations of that kind, why not just "explain" agents' actions in terms of an *action* table, which specifies what action the agent undertakes in every situation they might encounter?

Elga and Rayo acknowledge the difficulty, and seek to address it by saying that "An access table is only explanatory when its elicitation conditions are construed sufficiently broadly." (Elga and

Rayo 2019, p. 7). But even with broad elicitation conditions, the account does not vindicate the robustness of ordinary belief-based explanations of behaviour (see also Norby 2014). Suppose I ask you where the party is tomorrow night, and you tell me it will be on Rivington Street. What reason do you have to suppose I will act on this information tomorrow? I may believe it in my presently active fragment, but for all you know, and for all I know, a different fragment may activate tomorrow. And that fragment may not contain this information. Such a shift can happen even if I just have two belief fragments, both with broad elicitation conditions.

Are our belief-based explanations secretly buttressed by tacit assumptions about the circumstances under which the relevant beliefs are active? It would be curious that assumptions about elicitation conditions are never be made explicit, if they they play the pervasive role in cognition that Elga and Rayo say they do. It is also not clear how we are supposed to figure out the elicitation conditions of the beliefs we attribute to people. Suppose Will in fact plays the trump card. What does that tell us about the elicitation conditions of the belief $\mathbf{B}_T$? Nothing besides the fact that the condition was apparently met on this particular occasion. Any feature of the occasion could be responsible. Maybe $\mathbf{B}_T$ was triggered because Will has exactly five cards left, or because diamonds are the leading suit, or because of the time of day, or because someone just used the word "subcutaneous" in conversation. Elga and Rayo do not constrain the nature of these elicitation conditions in any way. Absent such constraints, the theory leaves many of the belief-based inferences one would ordinarily make about Will radically unjustified.

Yalcin's fragmentation theory faces an exactly parallel issue, but the involvement of questions does mitigate the severity of the problem. Suppose Will has the following views:

- The view F, ***Fred and I*** *have diamonds* in response to *Who has diamonds?*
- The view T, *I have* ***only spades and hearts*** in response to *What suits do I have?*
- The view L, *The bottom card is a* ***five of hearts or diamonds*** in response to *What is my bottom card?*
- The view G, ***Addis Ababa*** *is the capital of Ethiopia* in response to *What is the capital of Ethiopia?*

Much as before, the first three views each suggest a different course of action (respectively to F*ollow suit*, to *play the* T*rump card* and to L*ook at the cards*), while G suggests no clear course of action at all. But this time, it seems somehow easier to predict which of these beliefs is more likely to steer the course on this particular occasion.

In his papers, Yalcin is not really concerned with the relationship between belief and action, so he does not address this difficulty. But if we employ the notion of facing a question developed above, we can say something on his behalf. It would be bizarre, for instance, for Will to look to his views about Ethiopia to guide him on this particular occasion. There is nothing that map could say that would address his dilemma. As I said before, in an Elga/Rayo atlas, every map is a map of the whole world. That means that *a priori*, any map in the atlas could in principle address your current navigation problem. By contrast, in a Yalcin atlas, different maps depict different regions. Different maps are therefore inherently appropriate for different occasions. So Yalcin's theory suggests a natural explanation for why view G is not used on this occasion.

Will faces the question *Do I have diamonds*. Which of the three remaining maps should he consult to address this question? Should he ask *What is my bottom card*, *Who has diamonds* or *What suits do*

I have? Only answers to the latter two questions are bound to settle the issue, so those are the more natural views to look at. Thus we can rule out the view L: the issue *what Will's bottom card is* could have a bearing on his decision, but it is clearly less salient for this particular choice.

So only F and T are left. Unfortunately that is still not enough to yield a prediction about what Will is going to do. For F and T pull different ways: on the basis of F you would expect Will to follow suit. But on the basis of T you would expect him to play the trump card. He cannot do both, so both views cannot possibly be "on" — even though both are relevant on this occasion. For this reason, it looks like any systematic account tying Yalcin-style belief states to choices is still going to have to invoke elicitation conditions or something of the kind. For the same reasons as before, this is bound to undermine the force of ordinary belief-based inferences.

Inquisitive decision theory has no need for elicitation conditions. This is thanks to the constraint that agents' outright and probabilisitic views must agree on overlapping questions. That means that whenever two views bear on the same decision situation, they will give the agent matching guidance on that choice, even in cases where the views are inconsistent (§2.11). In particular, a conflict like the one between F and T never arises for an inquisitive agent. The questions *Who has diamonds?* and *What suits do I have?* intersect. Consequently, for any inquisitive agent with views on both questions, the views will agree on the overlap: *Do I have diamonds or not?* This constraint gels with our pre-theoretic judgments in this case: it is intuitively plausible that the views F and T should form an impossible combination.

I should clarify that inquisitive decision theory does not preclude the possibility of a divided mind in pathological cases, such as those involving split brain patients or people with delusions or dissociative identity disorder (Nagel 1971, Davies and Egan 2013). The inquisitive theorist is free to speculate that in such cases, multiple subagents co-exist within a single body, each with their own, separate inquisitive belief state. In fact, I am willing to go further than that. I am sympathetic to the thought that even in ordinary, healthy human beings there may be subagents whose beliefs are not identical to the beliefs of the organism of the whole. For instance, Tamar Gendler's cases of *alief* suggest that a person's intellectual faculties and their lower-level affective response systems can sometimes have directly opposing views (Gendler 2008). In the examples she gives, neither system has full control over the agents' actions, and so it seems like the person as a whole need not share the beliefs of either subagent. Aristotle's parts of the soul are a classical precedent for this sort of idea.

What I find misguided is the attempt to equate such phenomena with ordinary cases of memory and deductive failure. I think our natural inclination to posit a divided mind in these instances is due precisely to the fact that the attempt to provide an ordinary, unified belief-based explanation for the agent's behaviour fails. It is instructive to reflect on the pathological cases, because they help clarify what true belief fragmentation is like. A telling characteristic that sets them apart is that there are always particular, physical actions that can be performed by one of the subagents but not the other. In split brain patients, the left hemisphere controls the right hand and the right hemisphere controls the left. With dissociative identity patients, the presenting identity usually has exclusive control over the person's motor and verbal behaviour. And in the "alief" cases, only the intellectual part exercises direct control over verbal behaviour, and only the affective system has the ability to make the person tremble or sweat.

One important reason not to classify the inquisitive account of belief and credence as a fragmentation view is that we ought to reserve the term "fragmentation" for abnormal cases that exhibit this characteristic kind of disunity. These are the cases to which the term applies in the first instance, and to which it is best fitted. Insofar as this is a verbal dispute, it strikes me as a consequential one, because there is a value to having a word for that distinctive psychological phenomenon. Cases where a doxastic inconsistency is due to a failure to see the consequences of one's beliefs just do not belong to this category.

For instance, in MITTEN STATE MURDERS II, there is no evidence of this kind of disunity in Mandy's behaviour. It does not matter how Mandy is confronted with the question *How many murders happened in Detroit*. Whether it happens in conversation, or written down, or in some non-verbal context, Mandy's response will be to perform whatever action makes sense given her view that *there were **120***, whether that be raising her left hand, raising her right hand, or giving some sort of verbal or written reply. This disposition is also stable over time (no less stable than the doxastic state itself, anyway). The same observation holds true for her view of the Michigan murder figure. Both views can in principle affect any action of Mandy in any domain. That is why we attribute both of these beliefs to *Mandy* and not to some subagent: they unify all of Mandy's behavioural dispositions.

The contrast with fragmentation theories brings out what remained constant between inquisitive decision theory and the classical theory it is based on: namely the principle that all of an agent's actions should make sense given the totality of their beliefs. The fragmentation picture arises when you start fiddling with this, making some of an agent's choices answerable to one subset of their beliefs, while admitting that other decisions only make sense relative to a different subset. This idea radically undermines the unifying power of ordinary belief-based explanations: even an agent with only two fragments cannot be relied upon to act tomorrow on what they learnt today. Thankfully, we do not have to make that painful sacrifice to avoid the distorting idealisations of the classical theory.

Chapter 4.
Representation Theorems

In Chapter 3, I mentioned that all of classical and inquisitive decision theory can in a sense be seen to follow from the classical and inquisitive views of individual beliefs respectively. To be a bit more precise, if an agent has inquisitive (classical) beliefs at all, it follows that they have a full inquisitive (classical) belief and credence state, and act on those beliefs and credences. In this chapter, I state and prove the formal results that underpin those claims. This takes the form of three representation theorems that can ground the attribution of doxastic states not only to highly idealised classical agents, but also to agents who behave in the sorts of non-classical ways that were highlighted in previous chapters.

I should warn any unwary readers that after some leisurely sections at the beginning, this chapter turns very technical towards the end. Since the concepts introduced will not be used in chapter 5, the reader could happily skip §4.3-6 on a first reading. Indeed, on any reading, the proofs in §4.6 will mostly hold appeal to those who, to use Stephen Yablo's phrase, "are possessed by a powerful urge" (Yablo 2014, p. xi).

§4.1-2 explains the content of the three representation theorems, and explains why a more broadly applicable alternative to the classical representation theorems is desirable. §4.3-4 set up a formal framework for describing behavioural dispositions that can apply to any agent, no matter how irrational and erratic their behaviour. If we identify classical and inquisitive belief and credence with particular behavioural dispositions in this framework, this allows us to state the minimal rationality conditions on behaviour that are required for being a classical or an inquisitive agent. §4.5 specifies the dispositions associated with which classical and inquisitive doxastic states, and states the corresponding minimal rationality conditions. §4.6 supplies proofs for all the claims in §4.5.

4.1 Representation for Econs

The formal results below concern the conditions an agent's behaviour has to meet in order for classical or inquisitive decision theory to correctly characterise that behaviour. Or, on a different way of viewing the matter, you might say that these are the conditions under which it is appropriate to represent the agent as having a classical or an inquisitive doxastic state: hence the term "representation theorems".

These requirements on behaviour are described in a way that does not itself make reference to the agent's doxastic attitudes (except perhaps indirectly, as noted in §1.7). Unlike Savage's famous representation theorems, the results below are only about ascribing beliefs and credences. They say nothing about the conditions under which we can ascribe utilities to agents. In the present formalism, the utilities are incorporated in the definition of a decision problem, and are taken as given in describing the agent's behaviour.

Ramsey's representation theorem of 1926 is of this same broad type: he also takes the value of the rewards as a given, and then considers the circumstances under which credences can be ascribed.[25] Von Neumann and Morgenstern 1953, by contrast, pursued the opposite project: they took credences/probabilities as given in their axioms and then provided conditions under which a *utility* can be ascribed.

For the classical case, we will prove the following representation theorem:

> **Classical Representation Theorem.** An agent is classical if and only if they discern dominant options. (4.12)

The technical vocabulary used here will all be defined more precisely below, but roughly speaking, an agent is *classical* just in case they maximise expected utility with respect to a particular classical probability. The theorem describes the conditions under which this is true: the agent must "discern dominant options". This basically means that they will never make a series of bets that will lead to a guaranteed loss. In other words: discerning dominant options is being generally disposed to avoid *Dutch books*. One could paraphrase (4.12) as follows: *An agent may be represented as guided by classical credences just in case they always avoid Dutch books*. This makes theorem (4.12) very similar to the results often adduced in support of Bayesianism (compare for instance Ramsey 1926, Lehman 1955 and Hájek 2005).

---

[25] "Let us call the things a person ultimately desires 'goods' and let us at first assume that they are numerically measurable and additive." (Ramsey 1931 [1926], p. 173-4)

One reason to be interested in representation theorems is that they provide a way of grounding the attribution of doxastic states to agents. In particular, behaviouralists and functionalists look to characterise mental states in terms of behavioural dispositions. On the face of it, (4.12) gives them precisely what they need: it associates a precise set of behavioural dispositions with every possible classical credence function — namely the disposition of maximising expected value with respect to that probability. If we are prepared to leave the justification for attributing a utility aside for now, this immediately yields a way to characterise credences that is acceptable to behaviourists: an agent's credence state is just whichever classical probability the agent is disposed to maximise expected utility with respect to.

The trouble with this idea is that agents in the real world are not disposed to maximise expected utility with respect to any classical probability. As we saw in Chapters 1, 2 and 3, actual agents do not have classical behavioural dispositions at all. The reason, as anyone familiar with the literature on Dutch books is aware, is that complete immunity to Dutch books is an extremely high bar to clear: Dutch bookies are a crafty lot. Because of this, it does not look like (4.12) can ground the attribution of classical credences to ordinary agents after all.

A colourful way to make this point is to say that the classical representation theorem gives us a way to ground the attribution of credences to *econs* but not to humans. Econs are a fictional species who, unlike humans, behave in accordance with the laws of classical decision theory (see Thaler 2015, cf. Aumann 1987).[26] Of course, there are certain moves that can be made in an

[26] The Latin binomial for this species is *homo economicus*, or alternatively *homo rationalis*. A close cousin of the Econ is the *Logon*, a fictional believer whose doxastic and epistemic states obey the axioms of standard epistemic logic (Solaki, Berto and Smets 2019).

attempt to rescue this project. Sure, we may not be exactly like the econs, but at least it seems that we at some level aspire to achieve their level of coherence. Perhaps that makes us similar enough to econs to justify attributing classical credences to us "by extension" so to speak.

I do not object to this move in principle. When theorising, a certain amount idealisation is usually inevitable, and psychology is no exception to the rule. But I do think there are limits to how far it can be taken. For on the face of it, we are not similar to econs at all. Econs are perfectly consistent masterminds with unbounded computational resources. An econ can solve a Rubik's cube or factorise a twenty-digit integer in an instant. They are so clever they never have a need to think something through or deliberate. They instantly bring all relevant knowledge to bear on any practical issue they confront. They never have the experience of failing to reproduce a word or name that is on the tip of their tongue.

Human beings are not like that at all. And even if we aspire to the ideal that econs represent, our computational limitations mean that it is very far beyond our grasp. Moreover, it appears that those same limitations engender fundamental, structural dissimilarities between human and econ beliefs. In particular, humans have a lot of inconsistent beliefs, and we fail to see many of the consequences of our beliefs. So why suppose that human doxastic states resemble the doxastic states of econs at all?

An analogous complaint can be lodged about all the other classical representation theorems, like Ramsey 1926, Von Neumann and Morgenstern 1953, Savage 1972 and Bolker 1966/Bradley 1998/Joyce 1999. All of these theorems assume behaviour is constrained in ways that, by these authors' own admission, fail to apply to real-world behaviour. If the axioms of all these theories described real-life human or animal behaviour, the import and application of these results would be less mysterious. But as it stands, it is difficult to see just what the significance of these results is supposed to be.

Of course it is no coincidence that all those different axiomatic systems have to make similarly unrealistic assumptions. Save a few subtle differences, they are basically ways of describing the same classical pattern of betting behaviour from slightly different angles. And it does not matter what axioms you use to describe this behaviour: as long as the axioms succeed in capturing classical, expected-utility-maximising behaviour, they must be just as far removed from reality as the classical behaviour itself.

4.2 Representation for the Rest of Us

If we want to overcome this shortcoming of the classical theorem, and prove a result that can ground the attribution of attitudes to a wider class of agents, we need to start by expanding the space of potential agential attitudes beyond classical probability/utility pairs: all the classical states are already matched to possible Econ behaviours. Once the new attitudes are associated with behavioural dispositions in a precise way, we can investigate the conditions under which we can attribute an attitude from the broader class. This will then yield a generalisation of the classical representation theorem. Since there are more attitudes to go around, these conditions are bound to be weaker, and thus the theorem applies to a wider class of possible agents. The hope is that this wider class will include agents that are more like ourselves: representation for the rest of us. Notably, formal results of this kind have been established for non-additive (Choquet) theories of credence (Gilboa 1987 and Schmeidler 1989), and also for "cumulative" prospect theory (Tversky and Kahneman 1990; for more see Elliott 2015, §2.4, App. B). In this chapter, we will establish such a result for inquisitive decision theory.

In previous chapters, I have tried to show that the application of inquisitive decision theory to ordinary agents requires a great deal less idealisation and introduces much less distortion than classical decision theory does (even if it still requires some measure of idealisation). The inquisitive representation theorem is a reflection of this:

> **Inquisitive Representation Theorem.** An agent is inquisitive if and only if they discern better outcomes. (4.18)

Again, this new terminology will be defined more sharply below, but roughly speaking, an agent is inquisitive just in case they maximise expected utility with respect to an inquisitive probability. To discern better outcomes means to pick the superior option when faced with two *constant* choices, which comes down to a choice between outcomes (a constant choice is a choice that would yield the same outcome relative to every possible world). For instance, when given a choice between a thousand and two thousand dollars, an agent who discerns better outcomes will pick two thousand dollars, assuming they want to maximise monetary gain. Given a choice between burning a hundred dollar bill and not doing so, they pick the latter. *Failing* to discern better outcomes would be to wilfully select the raspberry ice-cream over mango ice-cream, with the goal of getting the mango ice-cream.

As these examples bring out, discerning better outcomes is a weak rationality constraint. It is nothing like avoiding Dutch bookies. One could even go so far as to argue that it is not really a constraint at all, but something that agents analytically have to satisfy. After all, is it not part of what we mean by saying that your goal is mango ice-cream that you do not wilfully avoid mango ice-cream, but get it when it is on offer? Here is a different way to make this point. Formally, the theory represents outcomes by their utility values. And you might think that if an agent's utility assigns one outcome a higher value than another, that just entails that the agent is disposed to pick the higher-valued outcome. This is part of what it means for one outcome to have higher utility than the other. On this view, the condition of discerning better outcomes is trivially satisfied, which means (4.18) in fact tells us that every agent, no matter how irrational, is an inquisitive agent.

Personally, I think that would be over-egging the cake a bit. Even if it is granted that there is some analytic connection between an outcome's having high utility and the disposition to pick that outcome, it still seems implausible that it is an analytic truth that our preferences between outcomes are representable by a utility function at all. An irrational person might have circular preferences, for instance. Or the proportionality could be off somehow: someone might for example find A and B to be twice as desirable as C, while also finding A to be twice as desirable as B. However, it is true that the assumption that an agent discerns better outcomes does not go beyond the assumption that in the context of any decision situation there is a real-valued utility function that faithfully represent the agent's conative attitudes towards the outcoms. That is a pretty basic assumption of the theoretical framework we have been using, and one with a good pedigree — it exactly mirrors Ramsey's assumption cited above. While this assumption is not empty, it is very weak next to the classical assumption of immunity to Dutch books. Moreover, it is a simplifying assumption about *desire,* not belief, and therefore much less worrisome for present purposes.

That means that the inquisitive representation theorem applies not only to econs but to a vastly larger class of agents with potentially far less systematic behavioural dispositions. As I have argued in previous chapters, these inquisitive agents include creatures much more like ourselves, who are terrible at Rubik's cubes and prime factorisation, and for whom deduction and memory retrieval take effort. So from the perspective of the project of leveraging a representation theorem to ground the attribution of credences to ordinary people, the inquisitive representation theorem seems like a much more promising starting point than the classical one.

At this point, someone might cry foul: how can I claim, on the one hand, that inquisitive decision theory puts barely any constraints on behavioural dispositions, while also maintaining, on the other hand, that the theory explains and predicts our behaviour? Surely the only way the theory can give a substantive explanation of behaviour is by making some claims that constrain this behaviour. You can't get something out of nothing! Has there been a sleight of hand?

Well, not exactly. If an agent discerns better outcomes, that only guarantees that they must have *some* inquisitive beliefs. At a minimum, they must believe the trivial quizposition $\{\mathscr{W}\}^{\{\mathscr{W}\}}$ — Yablo calls this quizposition *Whatever*. Suppose that *Whatever* is the only belief the agent has. Well, then inquisitive decision theory makes no predictions about their behaviour beyond the fact that they will pick the better outcome when given a choice. So it is true that you do not get anything out of nothing.

Inquisitive decision theory only begins generating more substantial predictions and explanations once we look at agents who have beliefs about non-trivial questions. In fact, as described in §3.4 above, the more questions an agent has answers to, the more predictable their behaviour becomes. But at the same time, every question corresponds to an additional coherence constraint on the agent's behavioural dispositions. The more questions an agent answers, the more rational and cohesive the agent's beliefs and actions become (compare §2.5-6). In the limit, if an agent is so coherent that they have views on every question, their doxastic state determines their behaviour in every decision situation, except when there is a tie in the expected utilities. Only econs meet that threshold. To apply inquisitive decision theory to a human agents, we must make the plausible assumption that, to a good approximation, their behavioural dispositions lie somewhere between those of the completely chaotic, unpredictable *Whatever* agent and the perfectly coherent econ.

The following result captures this broader picture:

> **General Representation Theorem.** An agent maximises inquisitive expected utility when faced with a question Q if and only if they discern dominant Q-options. (4.20)

Let's say a *Dutch Q-book* is a Dutch book with the feature that if you conjoin the questions raised by each bet in the series, that conjunction is part of Q. Informally, for an agent to discern dominant Q-options is for them to avoid all Dutch Q-books. Especially for small questions Q, avoiding Q-books is a great deal easier than avoiding Dutch books in general. That is because the restriction to Q-books significantly restricts the bag of tricks a Dutch bookie has at their disposal. For example, the only way you can get Dutch P-booked on a polar question P is to bet both for and against one and the same proposition. As the question Q gets bigger, the Q-books get trickier and the condition of avoiding them becomes increasingly demanding.

Both the classical and inquisitive representation theorems are easy consequences of the general representation theorem. An agent who discerns dominant options avoids Q-books for *all* questions Q. So they have credences about *every* question and maximise expected utility no matter what question they are faced with: these are econs or classical agents. Agents who only discern better outcomes discern dominant $\{\mathscr{W}\}$-options, and so they have inquisitive credences about at least one question and maximise expected utility when confronted with any question they have views about: they are inquisitive agents.

The picture evoked by the general representation theorem is that of a spectrum of increasingly "discerning" agents. The more dominant options they discern, the more rational they are, and the more predictable their behaviour. At one extreme, there are agents who do not avoid Dutch Q-books for any Q. Such agents are so unpredictable and erratic that they cannot even be assigned an inquisitive credence. At the other extreme, there are agents who avoid *all* Dutch books — they have an inquisitive credence that is defined on every question, which is just a classical credence: those are the perfectly systematic econs. And the rest of us? Well, we are somewhere in the middle.

4.3 Compound Decisions

We will be interested in situations where agents have to make a number of choices, either simultaneously or in short succession, where both the actions and the outcomes of the individual choices are independent of one another. What makes those situations interesting is that the agent can be viewed either as making two component choices or a single overarching compound choice. In §2.2 and §2.11 above, we already made reference to such compound choices, but for present purposes we will need a more systematic treatment.

As an example, consider Julie, who has to decide what clothes to put on for her long mid-April hike in the English countryside. She has two decisions to make: whether to bring her hat or not, and whether to bring a big warm coat or a lighter jacket. She really likes wearing her hat, but she would prefer for it not to get wet. And she prefers a big warm coat over a light jacket just in case it's going to be cold. Below, table 8 and 9 on the left represent these two decision problems, with p representing the possibility of pouring rain, and q the possibility that it will be quite cold. Table 10 on the right presents the same decision situation a different way, as a choice between four different outfits whose payoff depends on the overall weather conditions.

	p	¬p
k	1	4
l	3	2

	q	¬q
m	4	2
n	2	4

	pq	p¬q	¬pq	¬p¬q
k + **m**	5	3	8	6
k + **n**	3	5	6	8
l + **m**	7	5	6	4
l + **n**	5	7	4	6

TABLES 8-10: A COMPOUND DECISION

This compound decision problem can be computed from the smaller problems using a simple operation:

The ***sum*** of two decision problems Γ and Δ is the following the decision problem:

$$\Gamma + \Delta \quad =_{\text{df}} \quad \{\, \mathbf{a} + \mathbf{b} \ :\ \mathbf{a} \in \Gamma, \mathbf{b} \in \Delta \,\}$$

where "$\mathbf{a} + \mathbf{b}$" denotes the option $\mathbf{c}$ such that $\mathbf{c}(w) = \mathbf{a}(w) + \mathbf{b}(w)$ for all $w \in \mathscr{W}$.

(4.1)

Julie will make the decision { **k**, **l** } + { **m**, **n** } in virtue of making decisions { **k**, **l** } and { **m**, **n** } simultaneously. But it is important to stress that is *not* in general true that deciding Γ and also deciding Δ entails deciding Γ + Δ. If that were so, there would be no difference between a bet and a repeated bet: deciding { **x**, **y** } would amount to deciding { **x**, **y** } + { **x**, **y** } = { **x** + **x**, **x** + **y**, **y** + **y** }. Julie faces the sum of { **k**, **l** } and { **m**, **n** } because her two decisions are completely *independent* of one another: the outcome of each choice does not affect the outcome of the other, and what she chooses in one problem does not constrain what she chooses in the other. Only if an agent simultaneously makes two independent choices Γ and Δ does it follow that they are also making the choice represented by the sum of these problems Γ + Δ.

Conversely, deciding the sum Γ + Δ also does not entail deciding Γ or Δ. Deciding Γ and Δ simultaneously and independently is one way to decide Γ + Δ, but it is not the only way. For instance, Becky the financier faces the problem { **k**, **l** } + { **m**, **n** } displayed in in table 10 if she has a forced choice between four complex agricultural investments I, II, III and IV, whose respective financial returns she knows to depend on the April weather conditions in the manner indicated in the table. But Becky's choice is not, it seems, composed of a choice between *making investment*

I or II and *investment III or IV* on the one hand, and a choice between *making investment I or III* and *making investment II or IV* on the other.

For one, that unnatural division presupposes a plenitudinous metaphysics of actions with an unusual tolerance for disjunctive actions. But even if we were to grant that the choice can be split into these two strange components, whether the first of these is an instance of { **k**, **l** } still depends on how the credit for the overall profits is to be split between the component choices. And it turns out that no matter how we answer that question, there are counterexamples to the thesis that deciding Γ + Δ entails deciding Γ.

Let us say for instance that Becky chooses investment III and gets a return of seven grand, because it rains and is quite cold. How much of that return is due to her choice of *investment III or IV* over *investment I or II* and how much of it is due to her choice of *investment I or III* over *investment II or IV*? Suppose for the sake of argument that there is a principled answer to that question. I will even grant the division turns out to be three grand to the former and four grand to the latter. In fact let us grant that all entries in the table are split in such a way that the first choice does instantiate the decision problem { **k**, **l** }. Then there would still be a counterexample to the thesis that facing Γ + Δ entails facing Γ. For note that

$$\{ \mathbf{k}, \mathbf{l} \} + \{ \mathbf{m}, \mathbf{n} \} = \{ \mathbf{k} + 1, \mathbf{l} + 1 \} + \{ \mathbf{m} - 1, \mathbf{n} - 1 \}$$

But given what we've just granted, Becky does *not* face { **k** + 1, **l** + 1 } or { **m** – 1, **n** – 1 } in spite of facing their sum. And that is not the only counterexample: take any option **x**, and the following holds true: { **k**, **l** } + { **m**, **n** } = { **k** + **x**, **l** + **x** } + { **m** – **x**, **n** – **x** }.

4.4 Agential States

We also need a formal way of capturing the way an agent is disposed to respond to decision problems. The basic approach we will take here is to codify an agent's dispositions as a function that can take in any decision problem as an input and returns the way the agent is disposed to respond to that problem as its output. In principle, any agent whatsoever should have a dispositional profile of this kind, no matter how irrational and erratic their behaviour might be.

You might have thought that this function will represent an agent's response to a problem Δ as a particular option in Δ: whatever option the agent is disposed to choose in the situation Δ. But on reflection, that is too simplistic. There need not be any one option the agent is disposed to choose. For instance, suppose a quiz show participant is choosing between a blue door and a red door, knowing that the prize is behind one of the doors. If they consider one door more probable than the other one, they will presumably be disposed to pick that door. But what if the agent considers both possibilities equally likely? Then they need not have a firm disposition to pick one over the other.

To capture this possibility, we will in general want to characterise an agent's disposition to respond to a given decision problem as a *set* of options, all of which an agent might, given their dispositions, choose when confronted with that problem. If an agent has no firm dispositions at all about how to respond to a problem Δ, the set in question will simply be all of Δ. Formally speaking, such a non-empty subset of the original decision problem is itself also a decision problem. This leads to the following notion of an agential state:

> An ***agential state*** is a function $\boldsymbol{\alpha}$ from decision problems to decision problems satisfying the following three constraints for all Γ, Δ:
>
> i) $\varnothing \subset \boldsymbol{\alpha}(\Delta) \subseteq \Delta$
>
> ii) If $\mathbf{a} \notin \boldsymbol{\alpha}(\Delta)$, then $\boldsymbol{\alpha}(\Delta) = \boldsymbol{\alpha}(\Delta \setminus \{\mathbf{a}\})$
>
> iii) $(\boldsymbol{\alpha}(\Gamma) + \boldsymbol{\alpha}(\Delta)) \subseteq \boldsymbol{\alpha}(\Gamma + \Delta)$
>
> The options in $\Delta \setminus \boldsymbol{\alpha}(\Delta)$ are the options the agent α ***rules out*** in Δ. The options in $\boldsymbol{\alpha}(\Delta)$ are that α ***leaves open*** in Δ.[27] (4.2)

When referring to an agent as "α", we will always call their agential state "$\boldsymbol{\alpha}$", using the same lower-case Greek letter, but bold-faced. This notation might have been confusing if not for the fact that, since the agential state captures all the information we need about an agent, we need not pay much heed to the distinction between agents and their agential state. The motivation for constraint (i) in (3.3) is pretty clear: the agent must choose some option or other, so there always has to be at least one option compatible with their behavioural dispositions. And an agent can never choose an option they do not have.

The intuition behind condition (ii) is that when you choose between options **a**, **b**, **c**, … **z**, where

---

[27] If you agree with the view considered at the start of §4.2, that it is analytic that agents prefer higher-utility outcomes, then that would motivate the addition of a fourth condition to this definition:

> iv) For all $x, y \in \mathbb{R}$, if $x < y$ then $\boldsymbol{x} \notin \boldsymbol{\alpha}(\{\boldsymbol{x}, \boldsymbol{y}\} \cup \Delta)$

Here $\boldsymbol{x}$ and $\boldsymbol{y}$ represent constant options with value x and y respectively (see §4.5.1). With this fourth condition in place, we would get a version of the inquisitive representation theorem according to which every agent whatsoever is an inquisitive agent.

you rule out option **a**, you are in effect choosing between the remaining options **b**, **c**, … **z**. Removing a dispreferred option cannot make a difference for the agent's choice, since in ruling out the option the agent removes it anyway. This is essentially a weakening of the principle of *Independence of Irrelevant Alternatives* (IIA), which says that adding new options cannot affect an agent's preferences with respect to the old options.

The reason to resist endorsing a stronger version IIA is to allow for the possibility that a plentitude of options can drown out an option that would stand out if only the field were less crowded. Savage (1976) provides a nice example to illustrate this this: "You cannot be confident of having composed the ten word telegram that suits you best, though the list of possibilities is finite and vast numbers of possibilities can be eliminated immediately; after your best efforts, someone may suggest a clear improvement that had not occurred to you."

What (iii) says is that any combination of an option the agent might choose in Γ and an option they might choose in Δ is also left open in the problem Γ + Δ. To see why this has to be true, let us return to the example of Julie above. While deciding { **k**, **l** } and { **m**, **n** } independently is not the *only* way to decide the problem { **k**, **l** } + { **m**, **n** }, it is one possible way to do it. And as the case of Julie illustrates, when the decision is made in that particular way, her decision between **k** and **l** and her decision between **m** and **n**, jointly make for a decision between **k** + **m**, **k** + **n**, **l** + **m** and **l** + **n**. So when making both both small decisions independently, all combinations of responses Julie leaves open in the former situations are ipso facto possible responses to the larger problem when it is decided in this manner.

For instance, suppose $\mathbf{k} + \mathbf{n} \in \gamma(\{\mathbf{k}, \mathbf{l}\}) + \gamma(\{\mathbf{m}, \mathbf{n}\})$, where γ is Julie's agential state. Then $\mathbf{k} \in \gamma(\{\mathbf{k}, \mathbf{l}\})$ and $\mathbf{n} \in \gamma(\{\mathbf{m}, \mathbf{n}\})$, which is to say that Julie leaves open $\mathbf{k}$ in the problem $\{\mathbf{k}, \mathbf{l}\}$ and leaves open $\mathbf{n}$ in the problem $\{\mathbf{m}, \mathbf{n}\}$. It follows that if she is making both choices simultaneously and independently, Julie might choose the option $\mathbf{k} + \mathbf{n}$ in response to this larger problem, and thus $\mathbf{k} + \mathbf{n} \in \gamma(\{\mathbf{k}, \mathbf{l}\} + \{\mathbf{m}, \mathbf{n}\})$. More generally, any combination of options an agent in state $\boldsymbol{\alpha}$ leaves open in Γ and Δ is also left open when facing both problems independently, and thus also left open in $\Gamma + \Delta$. Thus $(\boldsymbol{\alpha}(\Gamma) + \boldsymbol{\alpha}(\Delta)) \subseteq \boldsymbol{\alpha}(\Gamma + \Delta)$. (But one may object to the strong notion of "independent" choices used here; cf. the note on p. 172 below.)

The opposite inclusion $\boldsymbol{\alpha}(\Gamma) + \boldsymbol{\alpha}(\Delta) \supseteq \boldsymbol{\alpha}(\Gamma + \Delta)$ fails because, as discussed in the previous section, the problem $\Gamma + \Delta$ can be faced without facing Γ. Since the agent's dispositions in response to Γ are irrelevant in that case, they might, for example, choose an option $\mathbf{a} + \mathbf{b}$ although $\mathbf{a}$ is ruled out in Γ, in which case $\mathbf{a} + \mathbf{b} \in \boldsymbol{\alpha}(\Gamma + \Delta)$ but $\mathbf{a} + \mathbf{b} \notin \boldsymbol{\alpha}(\Gamma) + \boldsymbol{\alpha}(\Delta)$.

The notion of an agential state allows us define an agent's preferences in terms of their behavioural dispositions as follows:

> An agent α ***prefers*** an option $\mathbf{a}$ to an option $\mathbf{b}$, written $\mathbf{a} \succ_\alpha \mathbf{b}$, just in case $\mathbf{b} \notin \boldsymbol{\alpha}(\Delta)$ for all problems Δ that contain both $\mathbf{a}$ and $\mathbf{b}$. (4.3)

Preference thus defined is a strict partial order. Because $\mathbf{a} \succ_\alpha \mathbf{b}$ entails that $\mathbf{a}$ but not $\mathbf{b}$ is in $\boldsymbol{\alpha}(\{\mathbf{a}, \mathbf{b}\})$, it follows that $\succ_\alpha$ is irreflexive and asymmetric. For transitivity, note that by clause (ii) of df. (4.2) it follows from $\mathbf{a} \succ_\alpha \mathbf{b}$ that for any Δ, $\boldsymbol{\alpha}(\{\mathbf{a}, \mathbf{c}\} \cup \Delta) = \boldsymbol{\alpha}(\{\mathbf{a}, \mathbf{b}, \mathbf{c}\} \cup \Delta)$. So if furthermore $\mathbf{b} \succ_\alpha \mathbf{c}$, then $\boldsymbol{\alpha}(\{\mathbf{a}, \mathbf{c}\} \cup \Delta) = \boldsymbol{\alpha}(\{\mathbf{a}, \mathbf{b}, \mathbf{c}\} \cup \Delta) = \boldsymbol{\alpha}(\{\mathbf{a}, \mathbf{b}\} \cup \Delta) = \boldsymbol{\alpha}(\{\mathbf{a}\} \cup \Delta)$, whence $\mathbf{a} \succ_\alpha \mathbf{c}$.

Besides this categorical notion of preference, we will also want to appeal to a question-specific notion of preference:

> An agent α *Q-**prefers*** an option **a** to an option **b**, written $\mathbf{a} \succ^{Q}_{\alpha} \mathbf{b}$, just in case $\mathbf{b} \notin \boldsymbol{\alpha}(\Delta)$ for all Q-raising problems Δ that contain both **a** and **b**. (4.4)

Q-preference is also a strict partial order.

4.5 Definitions and Results

This section lists the remaining definitions needed to state and prove the representation theorems. I take these definitions to be naturally motivated by the discussion in the preceding sections and in Chapters 1, 2 and 3. I list them here without further comments or explanation.

4.5.1 Some Notation

- *Differences*. The difference between **a** and **b**, written $\mathbf{a} - \mathbf{b}$, is the option **c** such that for all $w \in \mathscr{W}$, $\mathbf{c}(w) = \mathbf{a}(w) - \mathbf{b}(w)$.
- *Constant options*: **o** is the zero constant option and **i** the unity constant option: that is $\mathbf{o}(w) = 0$ and $\mathbf{i}(w) = 1$ for all $w \in \mathscr{W}$. When using italics for real numbers, boldfaced italics are the corresponding constant options. That is , for any $x \in \mathbb{R}$, $\boldsymbol{x}(w) = x$ for all $w \in \mathscr{W}$.
- *Restriction*: for any option **a** and any proposition p, $\mathbf{a}_{\mathrm{p}}$ is the option $\mathbf{a}_{\mathrm{p}}(w) = \mathbf{a}(w)$ for all $w \in \mathrm{p}$, and $\mathbf{a}_{\mathrm{p}}(w) = 0$ for all $w \notin \mathrm{p}$. (In particular, $\mathbf{i}_{\mathrm{p}}(w) = 1$ for all $w \in \mathrm{p}$, and $\mathbf{i}_{\mathrm{p}}(w) = 0$ for all $w \notin \mathrm{p}$.)
- *Scalar multiples*: $r \cdot \mathbf{a}$ is the option **b** such that $\mathbf{b}(w) = r \cdot \mathbf{a}(w)$ for all $w \in \mathscr{W}$; also $-\mathbf{a} = -1 \cdot \mathbf{a}$.
- *Weak dominance*: $\mathbf{a} \geq \mathbf{b}$ just in case $\mathbf{a}(w) \geq \mathbf{b}(w)$ for all $w \in \mathscr{W}$.

4.5.2 Classical Belief and Credence

An option **a** ***(strictly) dominates*** an option **b**, written $\mathbf{a} > \mathbf{b}$, just in case $\mathbf{a}(w) > \mathbf{b}(w)$ for all $w \in \mathscr{W}$. An option **a** ***(strictly) p-dominates*** an option **b**, written $\mathbf{a} >_{\mathrm{p}} \mathbf{b}$, just in case $\mathbf{a}(w) > \mathbf{b}(w)$ for all $w \in \mathrm{p}$. (4.5)

An agent α ***classically believes*** p just in case they will always prefer a p-dominant option. That is to say, just in case for any options **a** and **b**, $\mathbf{a} \succ_\alpha \mathbf{b}$ whenever $\mathbf{a} >_{\mathrm{p}} \mathbf{b}$. The ***classical belief state*** of an agent α is the following set of propositions:

$$\mathbf{CB}_\alpha \;=_{\mathrm{df}}\; \{\, \mathrm{p} \in \mathscr{P}(\mathscr{W}) : \alpha \text{ classically believes p} \,\} \tag{4.6}$$

Result. For any agent α, if $\mathbf{CB}_\alpha \neq \varnothing$, then $\mathbf{CB}_\alpha$ is a consistent classical information state. (4.7)

Proof. See §2.2 above.

An agent α ***discerns dominant options*** if and only if they always prefer strictly dominant options; that is, if and only if $\mathbf{a} \succ_\alpha \mathbf{b}$ whenever $\mathbf{a} > \mathbf{b}$. (4.8)

If p is a proposition and x is a real number in the interval [0, 1], then an option **a** ***x-p-governs*** an option **b**, written $\mathbf{a} >^x_{\mathrm{p}} \mathbf{b}$, just in case for any $w \in \mathrm{p}$, and any $v \in \neg\mathrm{p}$:

$$x \cdot (\mathbf{a} - \mathbf{b})(w) \quad > \quad (1 - x) \cdot (\mathbf{b} - \mathbf{a})(v)$$

Or equivalently, just in case $x \cdot \min_{\mathrm{p}} (\mathbf{a} - \mathbf{b}) \;>\; (1 - x) \cdot \max_{\neg\mathrm{p}} (\mathbf{b} - \mathbf{a})$. (4.9)

An agent α has ***classical credence*** x in the proposition p just in case the agent always prefers x-p-governing options. That is, just in case $\mathbf{a} \succ_\alpha \mathbf{b}$ whenever $\mathbf{a} >^x_{\mathrm{p}} \mathbf{b}$. The ***classical credence state*** of an agent α is the (partial) function $\mathbf{CCr}_\alpha \colon \mathscr{P}(\mathscr{W}) \to [0, 1]$

such that $\mathbf{CCr}_\alpha(\mathrm{p}) = x$ if and only if α has credence x in the proposition p. (4.10)

An agent α is ***classical*** if and only if they satisfy these conditions:

i) α's classical belief state $\mathbf{CB}_\alpha$ is a consistent classical information state

ii) α's classical credence state $\mathbf{CCr}_\alpha$ is a classical probability

iii) α maximises the expected utility given $\mathbf{CCr}_\alpha$. That is, for any decision problem Δ,

$$\boldsymbol{\alpha}(\Delta) \quad \subseteq \quad \{\, \mathbf{a} \in \Delta : \text{for all } \mathbf{b} \in \Delta,\ \mathcal{E}_{\mathbf{CCr}_\alpha}(\mathbf{a}) \geq \mathcal{E}_{\mathbf{CCr}_\alpha}(\mathbf{b}) \,\} \tag{4.11}$$

Classical Representation Theorem. An agent is classical if and only if they discern dominant options. (4.12)

4.5.3 Inquisitive Belief and Credence

An agent α ***inquisitively believes*** A_Q just in case they always Q-prefer A-dominant options. That is to say, just in case for any options $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} \succ^{\mathrm{Q}}_{\alpha} \mathbf{b}$ whenever $\mathbf{a} >_{\mathrm{A}} \mathbf{b}$. The ***inquisitive belief state*** of α is the following set of quizpositions:

$$\mathbf{QB}_\alpha \quad =_{\mathrm{df}} \quad \{\, \mathrm{A}_\mathrm{Q} : \alpha \text{ inquisitively believes } \mathrm{A}_\mathrm{Q} \,\} \tag{4.13}$$

Result For any agent α, if $\mathbf{QB}_\alpha \neq \varnothing$, then $\mathbf{QB}_\alpha$ is a coherent inquisitive information state. (4.14)

Proof. See §2.11 above.

An agent α has ***inquisitive credence*** x in the quizposition A^Q just in case the agent always prefers x-A-governing Q-options. The ***inquisitive credence state*** of an agent

α is the partial function $\mathbf{QCr}_\alpha\colon \mathcal{Q}(\mathcal{D}) \rightarrow [0, 1]$ such that $\mathbf{QCr}_\alpha(\mathrm{A^Q}) = x$ if and only if α has inquisitive credence x in the quizposition $\mathrm{A^Q}$. (4.15)

An agent α ***discerns better outcomes*** if and only if they always prefer strictly better outcomes; that is, if and only if for any $x, y \in \mathbb{R}$, $\boldsymbol{x} \succ_\alpha \boldsymbol{y}$ whenever $x > y$. (4.16)

An agent α is ***inquistive*** if and only if they satisfy these conditions:

i) α's inquisitive belief state $\mathbf{QB}_\alpha$ is a coherent inquisitive information state

ii) α's inquisitive credence state $\mathbf{QCr}_\alpha$ is an inquisitive probability

iii) α maximises the expected utility given $\mathbf{QCr}_\alpha$. That is, for any decision problem Δ raising a question $\mathrm{Q} \in \mathcal{D}_{\mathbf{QCr}_\alpha}$,

$$\boldsymbol{\alpha}(\Delta) \quad \subseteq \quad \{\, \mathbf{a} \in \Delta : \text{for all } \mathbf{b} \in \Delta,\ \mathcal{E}_{\mathbf{QCr}_\alpha}(\mathbf{a}) \geq \mathcal{E}_{\mathbf{QCr}_\alpha}(\mathbf{b}) \,\} \qquad (4.17)$$

Inquisitive Representation Theorem. An agent is inquisitive if and only if they discern better outcomes. (4.18)

4.5.4 A General Representation Theorem

An agent ***discerns dominant*** Q***-options*** if and only if they prefer strictly dominant options when faced with Q; that is, if and only if $\mathbf{a} \succ_\alpha^{\mathrm{Q}} \mathbf{b}$ whenever $\mathbf{a} > \mathbf{b}$. (4.19)

General Representation Theorem. An agent maximises inquisitive expected utility when faced with a question Q if and only if they discern dominant Q-options. (4.20)

4.6 Proofs

In this section, I prove the three representation theorems (4.12), (4.18) and (4.20). Since the only person we talk about in this section is our friend α the arbitrary agent, we will avoid some clutter by omitting the subscripted αs.

4.6.1 Coherence of Credences

This subsection proves that classical credences cohere with one another. These are conditional results of the form "If α has such and such classical credences, then…". In the next section we will discharge all those antecedents, showing that any agent who discerns dominant options has credences about all propositions.

Lemma 4.21 If an agent α has any classical credences at all, then α discerns dominant options and has classical credence 1 in $\top$.

Proof. Suppose agent α has classical credence x in some proposition p or other. Note that the condition for $\top$-1-governance is that $(\mathbf{a} - \mathbf{b})(w) > 0$ for all $w \in \top$, which is equivalent to $\mathbf{a} > \mathbf{b}$. So all we need to show is that α prefers $\mathbf{a}$ over $\mathbf{b}$ whenever $\mathbf{a} > \mathbf{b}$. Suppose $\mathbf{a} > \mathbf{b}$. Then for any $w \in \mathrm{p}$, $v \in \neg\mathrm{p}$ (indeed for any w and v at all), $x \cdot (\mathbf{a} - \mathbf{b})(w) > 0 > (1 - x) \cdot (\mathbf{b} - \mathbf{a})(v)$. In other words, $\mathbf{a} >^{x}_{\mathrm{p}} \mathbf{b}$, so that $\mathbf{a} \succ \mathbf{b}$. ■

Lemma 4.22 Suppose an agent has classical credence x in p, and classical credence y in q, where p entails q, then $x \leq y$. Consequently, an agent has at most one credence in a given proposition.

Proof. Suppose α has classical credence x in p, where $x > 0$. Let q be a proposition entailed by p and let y be such that $0 \leq y < x$. We show the agent does not have credence y in q. If $y = 0$, let ε be an arbitrary positive value. If $y > 0$ let $\varepsilon := (x - y)/2y > 0$. Consider the following bet:

	p	¬p
b	$1 - x + \varepsilon$	$-x$
o	0	0

Note that for all $w \in$ p, $v \in \neg$p:

$$x \cdot (\mathbf{b} - \mathbf{o})(w) = x \cdot (1 - x + \varepsilon) > x \cdot (1 - x) = (1 - x) \cdot (\mathbf{o} - \mathbf{b})(v)$$

Hence **b** p-x-dominates **o**, so that $\mathbf{b} \succ \mathbf{o}$. Now note that since p entails q, there are no p-worlds outside q so that $\mathbf{b}(v) = -x$ for all $v \in \neg$q. Hence we have, for all $w \in$ q and $v \in \neg$q,

$$\begin{aligned} y \cdot (\mathbf{o} - \mathbf{b})(w) \; - \; (1 - y) \cdot (\mathbf{b} - \mathbf{o})(v) \; &\geq \; -y \cdot (1 - x + \varepsilon) + (1 - y) \cdot x \\ &= \; x - y \; - \; y\varepsilon \; = \; \begin{cases} x & \text{if } y = 0 \\ (x - y)/2 & \text{if } y > 0 \end{cases} \\ &> \; 0 \end{aligned}$$

So **o** q-y-dominates **b**, and yet $\mathbf{o} \nsucc \mathbf{b}$. So α does not have credence y in q. (It follows from this result that if an agent has credences x and y in p, $x \geq y \geq x$, whence $x = y$.) ■

Lemma 4.23 Suppose p, q are disjoint propositions, and suppose α has classical credences x, y and z in propositions p, q and (p ∨ q) respectively. Then $z = x + y$.

Proof. Let α, p, q, x, y and z be as specified. Now suppose the agent simultaneously faces the a bet on p and a bet on q of the following kind, where $\varepsilon > 0$:

	p	¬p
$\mathbf{p}^{-}$	$1 - x - \varepsilon$	$-x$
o	0	0

	q	¬q
$\mathbf{q}^{-}$	$1 - y - \varepsilon$	$-y$
o	0	0

Note that $\mathbf{o} >_{\mathrm{p}}^{x} \mathbf{p}^{-}$: for all $w \in \mathrm{p}$, $v \in \neg\mathrm{p}$

$$x \cdot (\mathbf{p}^{-} - \mathbf{o})(w) = x \cdot \mathbf{p}^{-}(w) = x \cdot (1 - x - \varepsilon) < x \cdot (1 - x) = (1 - x) \cdot (\mathbf{o} - \mathbf{p}^{-})(v)$$

Thus $\mathbf{o} \succ \mathbf{p}^{-}$, and likewise $\mathbf{o} \succ \mathbf{q}^{-}$. Hence $\boldsymbol{\alpha}(\{\, \mathbf{p}^{-}, \mathbf{o} \,\}) + \boldsymbol{\alpha}(\{\, \mathbf{q}^{-}, \mathbf{o} \,\}) = \{\, \mathbf{o} \,\} + \{\, \mathbf{o} \,\} = \{\, \mathbf{o} \,\}$. Hence by clause (4.2.iii) in the definition of an agential state, $\mathbf{o} \in \boldsymbol{\alpha}(\{\, \mathbf{p}^{-}, \mathbf{o} \,\} + \{\, \mathbf{q}^{-}, \mathbf{o} \,\})$. Here is a matrix for that problem $\{\, \mathbf{p}^{-}, \mathbf{o} \,\} + \{\, \mathbf{q}^{-}, \mathbf{o} \,\}$:

	p¬q	¬pq	¬(p ∨ q)
$\mathbf{p}^{-} + \mathbf{q}^{-}$	$1 - (x + y) - \varepsilon$	$1 - (x + y) - \varepsilon$	$-(x + y)$
$\mathbf{p}^{-}$	$1 - x - \varepsilon$	$-x$	$-x$
$\mathbf{q}^{-}$	$-y$	$1 - y - \varepsilon$	$-y$
$\mathbf{o}$	0	0	0

Note that there is no column pq because p and q are disjoint. Option $\mathbf{o}$ is left open in this problem, so apparently $\mathbf{p}^{-} + \mathbf{q}^{-} \nsucc \mathbf{o}$. Hence $\mathbf{p}^{-} + \mathbf{q}^{-}$ apparently does not (p ∨ q)-z-dominate $\mathbf{o}$. So it must be that for some $w \in (\mathrm{p} \vee \mathrm{q})$, $v \in \neg(\mathrm{p} \vee \mathrm{q})$,

$$\begin{aligned} z \cdot (\mathbf{p}^{-} + \mathbf{q}^{-})(w) &\leq (1 - z) \cdot -(\mathbf{p}^{-} + \mathbf{q}^{-})(v) \\ z \cdot (1 - (x + y) - \varepsilon) &\leq (1 - z) \cdot (x + y) \\ z \cdot (1 - \varepsilon) &\leq (x + y) \end{aligned}$$

Since this is true for all positive ε, it follows that $(x + y) \geq z$. For the other inequality, we run much the same argument. Consider of the following decision problems:

	p	¬p
$\mathbf{p}^{+}$	$1 - x + \varepsilon$	$-x$
$\mathbf{o}$	0	0

	q	¬q
$\mathbf{q}^{+}$	$1 - y + \varepsilon$	$-y$
$\mathbf{o}$	0	0

Here $\mathbf{p}^{+} >_{\mathrm{p}}^{x} \mathbf{o}$ and so that $\mathbf{p}^{+} \succ \mathbf{o}$, and $\mathbf{q}^{+} >_{\mathrm{q}}^{y} \mathbf{o}$ so that $\mathbf{q}^{+} \succ \mathbf{o}$. We conclude in the same way as before that $\mathbf{o} \nsucc \mathbf{p}^+ + \mathbf{q}^+$, and from that it follows that for some $w \in (\mathrm{p} \vee \mathrm{q})$, $v \in \neg(\mathrm{p} \vee \mathrm{q})$

$$\begin{aligned}
-z \cdot (\mathbf{p}^+ + \mathbf{q}^+)(w) &\leq (1-z) \cdot (\mathbf{p}^+ + \mathbf{q}^+)(v) \\
z \cdot (\mathbf{p}^+ + \mathbf{q}^+)(w) &\geq (1-z) \cdot -(\mathbf{p}^+ + \mathbf{q}^+)(v) \\
z \cdot (1 - (x+y) + \varepsilon) &\geq (1-z) \cdot (x+y) \\
z \cdot (1+\varepsilon) &\geq (x+y)
\end{aligned}$$

Since this is again true for arbitrarily small ε, that gives us $z \geq (x+y)$. ■

Theorem 4.24 Suppose α has classical credences about every proposition. Then α's classical credences form a classical probability $\mathbf{CCr}_\alpha$.

Proof. Lemma 4.22 ensures there is only one classical credence per proposition, so that we can define the function $\mathbf{CCr}_\alpha$. Lemma 4.21 give us that $\mathbf{CCr}_\alpha(\top) = 1$, and lemma 4.23 gives us finite additivity. ■

4.6.2 From $\mathcal{E}$-Values to Expected Values

This subsection demonstrates that agents who discern dominant options do in fact have credences about every proposition. Combined with theorem 4.24, this yields the conclusion that their credences are a classical probability.

Lemma 4.25 If $\mathbf{a} \succ \mathbf{b}$ then for any option $\mathbf{c}$, $\mathbf{c} + \mathbf{a} \succ \mathbf{c} + \mathbf{b}$. (In particular, $-\mathbf{b} \succ -\mathbf{a}$.)

We show the contrapositive. Suppose that for some $\mathbf{c}$, $\mathbf{c} + \mathbf{a} \nsucc \mathbf{c} + \mathbf{b}$, then by definition (4.3), there is some Δ such that $\mathbf{c} + \mathbf{b} \in \boldsymbol{\alpha}(\Delta \cup \{\mathbf{c} + \mathbf{a}, \mathbf{c} + \mathbf{b}\})$. Hence $\mathbf{b} \in \boldsymbol{\alpha}(\{-\mathbf{c}\}) + \boldsymbol{\alpha}(\Delta \cup \{\mathbf{c} + \mathbf{a}, \mathbf{c} + \mathbf{b}\})$. But note that $\{-\mathbf{c}\} + \Delta \cup \{\mathbf{c} + \mathbf{a}, \mathbf{c} + \mathbf{b}\} = \Gamma \cup \{\mathbf{a}, \mathbf{b}\}$ for some Γ. Hence by (4.2.iii), $\mathbf{b} \in \boldsymbol{\alpha}(\Gamma \cup \{\mathbf{a}, \mathbf{b}\})$, whence $\mathbf{a} \nsucc \mathbf{b}$. (To get the corollary that $\mathbf{a} \succ \mathbf{b}$ entails $-\mathbf{b} \succ -\mathbf{a}$, consider the case $\mathbf{c} = -(\mathbf{a} + \mathbf{b})$.) ■

Definition 4.26 For every action $\mathbf{a}$, the ***$\mathcal{E}$-value*** of $\mathbf{a}$ is $\mathcal{E}(\mathbf{a}) = \sup\{\, x : \mathbf{a} \succ \boldsymbol{x} \,\}$.

Note: As the notation foreshadows, we will prove below that the $\mathcal{E}$-value is identical to the classical expected value of the option for the agent. By standard mathematical convention, if $\{\, x : \mathbf{a} \succ \boldsymbol{x} \,\}$ is empty, $\mathcal{E}(\mathbf{a}) = -\infty$, and if it is unbounded above, $\mathcal{E}(\mathbf{a}) = +\infty$. But as long as α discerns dominant options, and $\mathbf{a}$ is bounded, neither case is instantiated. And assuming α discerns better outcomes, $\mathcal{E}(\boldsymbol{z}) = z$ for all $z \in \mathbb{R}$, since $\boldsymbol{x} \succ \boldsymbol{y}$ just in case $x > y$.

Lemma 4.27 If α discerns dominant options, then for any action $\mathbf{a}$, $\mathcal{E}(\mathbf{a}) = \inf\{\, x : \boldsymbol{x} \succ \mathbf{a} \,\}$

Proof: Assume α discerns dominant options. Then $\boldsymbol{x} \succ \boldsymbol{y}$ just in case $x > y$, so by transitivity of $\succ$, $\{\, x : \mathbf{a} \succ \boldsymbol{x} \,\}$ is closed downwards and $\{\, x : \boldsymbol{x} \succ \mathbf{a} \,\}$ is closed upwards. By antisymmetry of $\succ$, the sets are disjoint so that $\inf\{\, x : \boldsymbol{x} \succ \mathbf{a} \,\} \geq \sup\{\, x : \mathbf{a} \succ \boldsymbol{x} \,\} = \mathcal{E}(\mathbf{a})$. As just pointed out, since α discerns dominant options, we can suppose $\mathcal{E}(\mathbf{a}) < +\infty$. [28]

Now let r be any real number greater than $\mathcal{E}(\mathbf{a})$. Suppose for contradiction that $\boldsymbol{r} \nsucc \mathbf{a}$. Let s be any number between $\mathcal{E}(\mathbf{a})$ and r. It must be that $\mathbf{a} \nsucc \boldsymbol{s}$, because $s > \mathcal{E}(\mathbf{a}) = \sup\{\, x : \mathbf{a} \succ \boldsymbol{x} \,\}$. Because $\boldsymbol{r} \nsucc \mathbf{a}$ and $\mathbf{a} \nsucc \boldsymbol{s}$, there are some Γ and Δ such that $\mathbf{a} \in \boldsymbol{\alpha}(\Gamma \cup \{\boldsymbol{r}, \mathbf{a}\})$ and $\boldsymbol{s} \in \boldsymbol{\alpha}(\Delta \cup \{\boldsymbol{s}, \mathbf{a}\})$. Hence, by (4.2.iii), $\mathbf{a} + \boldsymbol{s} \in \boldsymbol{\alpha}(\Delta \cup \{\boldsymbol{r}, \mathbf{a}\}) + \boldsymbol{\alpha}(\Gamma \cup \{\boldsymbol{s}, \mathbf{a}\})$. By (4.2.iii), it follows from this that $\mathbf{a} + \boldsymbol{s} \in \boldsymbol{\alpha}(\Delta \cup \{\boldsymbol{r}, \mathbf{a}\} + \Gamma \cup \{\boldsymbol{s}, \mathbf{a}\})$. But then $\boldsymbol{r} + \mathbf{a} \nsucc \mathbf{a} + \boldsymbol{s}$ which contradicts the assumption that α discerns dominant options. So $\boldsymbol{r} \succ \mathbf{a}$ after all. So any real number r greater than $\mathcal{E}(\mathbf{a})$ is also greater than $\inf\{\, x : \boldsymbol{x} \succ \mathbf{a} \,\}$, whence $\mathcal{E}(\mathbf{a}) = \sup\{\, x : \mathbf{a} \succ \boldsymbol{x} \,\} = \inf\{\, x : \boldsymbol{x} \succ \mathbf{a} \,\}$. ■

---

[28] But even if $\mathbf{a}$ were unbounded, and $\mathcal{E}(\mathbf{a}) = +\infty$, the result would still hold. For then $\{\, x : \boldsymbol{x} \succ \mathbf{a} \,\} = \varnothing$ so that $\inf\{\, x : \boldsymbol{x} \succ \mathbf{a} \,\} = +\infty = \mathcal{E}(\mathbf{a})$

Corollary 4.28 If α discerns dominant options and $\mathcal{E}(\mathbf{a}) > \mathcal{E}(\mathbf{b})$, then $\mathbf{a} \succ \mathbf{b}$.

Suppose α discerns dominant options and $\mathcal{E}(\mathbf{a}) > \mathcal{E}(\mathbf{b})$. Then by Lemma 4.27, for any value y in the interval between them, $\mathbf{a} \succ \boldsymbol{y} \succ \mathbf{b}$. ■

Note: Lemma 4.27 is essentially involved in showing that credences must be *sharp*. It uses an insight from Adam Elga's (2010) argument against fuzzy credences. For most other results above, one can make do with a weaker version of (4.2.iii), namely

4.2.iii* For all Γ, Δ, and $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}(\Gamma + \Delta) \cap (\boldsymbol{\alpha}(\Gamma) + \boldsymbol{\alpha}(\Delta)) \neq \varnothing$

However, this proof makes essential use of the stronger (4.2.iii). Defenders of fuzzy credences may wish to consider adopting (4.2.iii*) instead, though I have not investigated this possibility.

Lemma 4.29 If α discerns dominant options, then $\mathcal{E}$ is linear: $\mathcal{E}(r \cdot \mathbf{a} + s \cdot \mathbf{b}) = r \cdot \mathcal{E}(\mathbf{a}) + s \cdot \mathcal{E}(\mathbf{b})$ for any options $\mathbf{a}$, $\mathbf{b}$, and any real numbers r and s. Furthermore, $\mathcal{E}(\mathbf{a}) \geq \mathcal{E}(\mathbf{b})$ whenever $\mathbf{a} \geq \mathbf{b}$.

Proof. We split up the result into the following sublemmas:

i) $\mathcal{E}(-\mathbf{a}) = -\mathcal{E}(\mathbf{a})$.

Using lemmas 4.25 and 4.27, $\mathcal{E}(-\mathbf{a}) = \inf\{ -x : -\boldsymbol{x} \succ -\mathbf{a} \} = \inf\{ -x : \mathbf{a} \succ \boldsymbol{x} \}$

$$= -\sup\{ x : \mathbf{a} \succ \boldsymbol{x} \} = -\mathcal{E}(\mathbf{a})$$

ii) $\mathcal{E}(\mathbf{a} + \mathbf{b}) = \mathcal{E}(\mathbf{a}) + \mathcal{E}(\mathbf{b})$.

Suppose $\mathbf{a} \succ \boldsymbol{x}$ and $\mathbf{b} \succ \boldsymbol{y}$. By lemma 4.25, $\mathbf{a} + \mathbf{b} \succ \mathbf{a} + \boldsymbol{y} \succ \boldsymbol{x} + \boldsymbol{y}$. So

$$\{ x + y : \mathbf{a} \succ \boldsymbol{x}, \mathbf{b} \succ \boldsymbol{y} \} \subseteq \{ z : \mathbf{a} + \mathbf{b} \succ \boldsymbol{z} \}.$$

Therefore, by definition of $\mathcal{E}$, $\mathcal{E}(\mathbf{a} + \mathbf{b}) \geq \mathcal{E}(\mathbf{a}) + \mathcal{E}(\mathbf{b})$. Likewise, $\mathcal{E}(\mathbf{a}) = \mathcal{E}(-\mathbf{b} + (\mathbf{a} + \mathbf{b})) \geq \mathcal{E}(-\mathbf{b}) + \mathcal{E}(\mathbf{a} + \mathbf{b}) = -\mathcal{E}(\mathbf{b}) + \mathcal{E}(\mathbf{a} + \mathbf{b})$ (using (i)). So $\mathcal{E}(\mathbf{a}) + \mathcal{E}(\mathbf{b}) \geq \mathcal{E}(\mathbf{a} + \mathbf{b})$.

iii) If $\mathbf{a} \geq \mathbf{b}$, $\mathcal{E}(\mathbf{a}) \geq \mathcal{E}(\mathbf{b})$

Suppose $\mathbf{a} \geq \mathbf{b}$. Then $\mathbf{a} - \mathbf{b} > \boldsymbol{x}$ for any negative x. So since α discerns dominant options, $\mathbf{a} - \mathbf{b} \succ \boldsymbol{x}$ for all $x < 0$. It follows by the definition of an $\mathcal{E}$-value (4.26) that $\mathcal{E}(\mathbf{a} - \mathbf{b}) \geq 0$, so using (ii), $\mathcal{E}(\mathbf{a}) = \mathcal{E}(\mathbf{b}) + \mathcal{E}(\mathbf{a} - \mathbf{b}) \geq \mathcal{E}(\mathbf{b})$.

iv) For any integer n, $\mathcal{E}(n \cdot \mathbf{a}) = n \cdot \mathcal{E}(\mathbf{a})$

$\mathcal{E}(0 \cdot \mathbf{a}) = \mathcal{E}(\mathbf{o}) = 0$. If n is positive, $\mathcal{E}(n \cdot \mathbf{a}) = \mathcal{E}(\mathbf{a} + \ldots + \mathbf{a}) = \mathcal{E}(\mathbf{a}) + \ldots + \mathcal{E}(\mathbf{a}) = n \cdot \mathcal{E}(\mathbf{a})$, using (ii). If n is negative, we get the result from (i).

v) For any rational number q, $\mathcal{E}(q \cdot \mathbf{a}) = q \cdot \mathcal{E}(\mathbf{a})$

Suppose $q = m/n$. By (iv), $n \cdot \mathcal{E}(q \cdot \mathbf{a}) = \mathcal{E}(nq \cdot \mathbf{a}) = \mathcal{E}(m \cdot \mathbf{a}) = m \cdot \mathcal{E}(\mathbf{a})$

vi) For any real number r, $\mathcal{E}(r \cdot \mathbf{a}) = r \cdot \mathcal{E}(\mathbf{a})$

Let $\mathrm{p} = \{w : \mathbf{a}(w) \geq 0\}$, so that $\mathbf{a}_\mathrm{p} \geq \mathbf{o}$. Note that $\mathbf{a} = \mathbf{a}_\mathrm{p} + \mathbf{a}_{\neg\mathrm{p}}$. Now let q, q' be any rationals such that $r \in (q, q')$. Then $q \cdot \mathbf{a}_\mathrm{p} \leq r \cdot \mathbf{a}_\mathrm{p} \leq q' \cdot \mathbf{a}_\mathrm{p}$. Hence $\mathcal{E}(q \cdot \mathbf{a}_\mathrm{p}) \leq \mathcal{E}(r \cdot \mathbf{a}_\mathrm{p}) \leq \mathcal{E}(q' \cdot \mathbf{a}_\mathrm{p})$, using (iii). So by (v), $q \cdot \mathcal{E}(\mathbf{a}) \leq \mathcal{E}(r \cdot \mathbf{a}_\mathrm{p}) \leq q' \cdot \mathcal{E}(\mathbf{a})$ for any such q, q'. So it must be that $\mathcal{E}(r \cdot \mathbf{a}_\mathrm{p}) = r \cdot \mathcal{E}(\mathbf{a}_\mathrm{p})$. Similarly, $\mathcal{E}(r \cdot \mathbf{a}_{\neg\mathrm{p}}) = r \cdot \mathcal{E}(\mathbf{a}_{\neg\mathrm{p}})$ so that $\mathcal{E}(r \cdot \mathbf{a}) = r \cdot (\mathcal{E}(\mathbf{a}_\mathrm{p}) + \mathcal{E}(\mathbf{a}_{\neg\mathrm{p}})) = r \cdot \mathcal{E}(\mathbf{a})$. ■

Theorem 4.30 If α discerns dominant options, and p is any proposition, then α has classical credence $\mathcal{E}(\mathbf{i}_\mathrm{p})$ in the proposition p. And α's credence state is the classical probability $\mathbf{CCr}_\alpha(\mathrm{p}) = \mathcal{E}(\mathbf{i}_\mathrm{p})$.

Proof. Let p be any proposition, let $c = \mathcal{E}(\mathbf{i}_\mathrm{p})$, and let $\mathbf{a}$ and $\mathbf{b}$ be any actions such that $\mathbf{a} >^c_\mathrm{p} \mathbf{b}$. We need to show that $\mathbf{a} \succ \mathbf{b}$. By the definition of governance (4.9), we have for all $w \in \mathrm{p}$, $v \in \neg\mathrm{p}$

$$c \cdot (\mathbf{a} - \mathbf{b})(w) \quad > \quad (1 - c) \cdot (\mathbf{b} - \mathbf{a})(v) \qquad (*)$$

Let $r = \min_\mathrm{p} \mathbf{a} - \mathbf{b}$ and $s = \min_{\neg\mathrm{p}} \mathbf{a} - \mathbf{b} = -\max_{\neg\mathrm{p}} \mathbf{b} - \mathbf{a}$. Then from (*) we get $c \cdot r > (1 - c) \cdot (-s)$, whence $c \cdot r + (1 - c) \cdot s > 0$.

Now consider the option $\mathbf{m} := r \cdot \mathbf{i}_{\mathrm{p}} + s \cdot \mathbf{i}_{\neg\mathrm{p}}$. Note $(\mathbf{a} - \mathbf{b}) \geq \mathbf{m}$, and also note that, by lemma 4.29, $\mathcal{E}(\mathbf{i}_{\neg\mathrm{p}}) = \mathcal{E}(\mathbf{i} - \mathbf{i}_{\mathrm{p}}) = \mathcal{E}(\mathbf{i}) - \mathcal{E}(\mathbf{i}_{\mathrm{p}}) = 1 - c$. Thus we have, using lemma 4.29 again, that

$$\begin{aligned}\mathcal{E}(\mathbf{a}) \; = \; \mathcal{E}(\mathbf{b}) + \mathcal{E}(\mathbf{a} - \mathbf{b}) \; \geq \; \mathcal{E}(\mathbf{b}) + \mathcal{E}(\mathbf{m}) \; &= \; \mathcal{E}(\mathbf{b}) \; + r \cdot \mathcal{E}(\mathbf{i}_{\mathrm{p}}) + s \cdot \mathcal{E}(\mathbf{i}_{\neg\mathrm{p}}) \\ &= \; \mathcal{E}(\mathbf{b}) \; + \; c \cdot r + (1 - c) \cdot s \; > \; \mathcal{E}(\mathbf{b})\end{aligned}$$

And hence by corollary 4.28, $\mathbf{a} \succ \mathbf{b}$, which is what we wanted to show. Thus $\mathcal{E}(\mathbf{i}_{\mathrm{p}})$ is α's classical credence in p. By theorem 4.24 above, the function $\mathbf{CCr}_\alpha(\mathrm{p}) = \mathcal{E}(\mathbf{i}_{\mathrm{p}})$ is a classical probability. ■

Corollary 4.31 If α discerns dominant options, $\mathbf{a} \succ \mathbf{b}$ whenever $\mathcal{E}_{\mathrm{CCr}_\alpha}(\mathbf{a}) > \mathcal{E}_{\mathrm{CCr}_\alpha}(\mathbf{b})$.

Proof. We need to show that for any $\mathbf{a}$, $\mathcal{E}(\mathbf{a}) = \mathcal{E}_{\mathrm{CCr}_\alpha}(\mathbf{a})$. The result will then immediately follow by corollary 4.28. For all the possible values u of $\mathbf{a}$, let $\mathbf{i}_u(w) = 1$ if $\mathbf{a}(w) = u$, and $\mathbf{i}_u(w) = 0$ otherwise, so that we can write $\mathbf{a} = \Sigma_u \; u \cdot \mathbf{i}_u$. By theorem 4.30, $\mathcal{E}(\mathbf{i}_u) = \mathbf{CCr}_\alpha\,(\{\, w : \mathbf{a}(w) = u \,\})$. Thus by linearity (4.29),

$$\mathcal{E}(\mathbf{a}) = \mathcal{E}(\Sigma_u \, u \cdot \mathbf{i}_u) = \Sigma_u \, u \cdot \mathcal{E}(\mathbf{u}) = \Sigma_u \, u \cdot \mathbf{CCr}_\alpha\,(\{\, w : \mathbf{a}(w) = u \,\}) = \mathcal{E}_{\mathrm{CCr}_\alpha}(\mathbf{a}). \;■$$

Note. This proof, as well as the definition (3.2) of a classical expected value, use the assumption that $\mathbf{a}$ has only finitely many possible values. This follows from the assumption that there are finitely many possible worlds, but both the definition of expected value and the result (4.31) can be extended to cover the case where $\mathbf{a}$ has infinitely many different outcomes. In that case, we will want to build it into the definition of an option that options are to be bounded functions. (This is to avoid paradoxes like those discussed in McGee 1999, Arntzenius and Barrett 1999 and Arntzenius, Elga and Hawthorne 2004.) Then we can approximate any option $\mathbf{a}$ below and above by a sequence of step functions with finite ranges that pointwise converge on $\mathbf{a}$. Using the fact that $\mathcal{E}(\mathbf{a}) \geq \mathcal{E}(\mathbf{b})$ whenever $\mathbf{a} \geq \mathbf{b}$ (sublemma 4.29.iii) a squeezing argument will then show that $\mathcal{E}(\mathbf{a}) = \mathcal{E}_{\mathrm{CCr}_\alpha}(\mathbf{a})$ whenever $\mathcal{E}_{\mathrm{CCr}_\alpha}(\mathbf{a})$ is defined.

Theorem 4.12 All and only agents who discern dominant options are classical

Proof. Result 4.21 shows that discerning dominant options is necessary for the satisfaction of clause (i) and (ii) of the definition of a classical agent (4.11). So classical agents discern dominant options. Conversely, theorem 4.30 shows that agents who discern dominant options satisfy clause (i) and (ii), and corollary 4.31 shows they satisfy clause (iii). ■

4.6.3 The Inquisitive Case

In this subsection, I prove the Inquisitive Representation Theorem (4.18) and the General Representation Theorem (4.20). The work done in the preceding subsections easily carries over to attitudes about a given question (lemma 4.32 below); all we need to establish is that the agent's attitudes to different questions are linked in the appropriate way.

Lemma 4.32 For any question Q, an agent has inquisitive beliefs and credences about Q only if α discerns Q-dominance. If α discerns Q-dominance, then:

A) α has inquisitive credence 1 in Q^{Q}.

B) For any $A \subseteq B \subseteq Q$, α's inquisitive credence in A^{Q} (if any) is less than or equal to their inquisitive credence in B^{Q} (if any). So α has at most one credence in a given quizposition.

C) For disjoint $A, B \subseteq Q$, α's inquisitive credence in $(A \cup B)^{Q}$ (if any) is the sum of their inquisitive credence in A^{Q} and in B^{Q} (if any).

D) For all $A \subseteq Q$, α has an inquisitive credence in A^{Q}

E) When faced with Q, α always maximises inquisitive expected utility with respect to those credences.

Proof. If we put $\mathscr{W} = Q$, the definitions classical belief and classical credence in $A \subseteq Q$ coincide exactly with those of inquisitive belief and inquisitive credence in the quizposition A^Q. Thus we get (A) from lemma 4.21, (B) from 4.22, (C) from 4.23, (D) from 4.30, and (E) from 4.31. ■

Corollary 4.33 If an agent has credences about some question Q, they have credences about every question that is part of Q.

Proof. If R is part of Q, discerning Q-dominance entails discerning R-dominance (from df. 4.19 above). ■

Lemma 4.34 An agent α has inquisitive credences about some quizpositions if and only if they discern better outcomes.

Proof. If α prefers better outcomes, then they have credence 1 in $\{ \mathscr{W} \}^{\{ \mathscr{W} \}}$ and credence 0 in $\bot^{\{ \mathscr{W} \}}$ (directly from the definitions 4.15 and 4.16). Conversely, if an agent has a credence in some quizposition A^Q, then they have credence 1 in Q^Q (from clause (A) of 4.32). It then follows from 4.15 that α discerns better outcomes. ■

Lemma 4.35 If quizpositions A^Q and B^R are intensionally equivalent, and some an agent has inquisitive credence x in A^Q and inquisitive credence y in B^R, then $x = y$.

Proof. Suppose α has inquisitive credence x in A^Q and y in B^R, and $p = \bigcup A = \bigcup B$. We show that x and y are both equal to α's inquisitive credence in the quizposition $\{ p \}^{\{p, \neg p\}}$. For let Δ be any decision situation that raises the polar question $\{ p, \neg p \}$, containing an option **a** that x-p-governs another option **b** in Δ. Then, since Δ also raises the bigger question Q (2.13), $\mathbf{b} \notin \boldsymbol{\alpha}(\Delta)$ because α has credence x in A^Q. So the agent $\{ p, \neg p \}$-prefers **a** to **b**. Thus they have credence x in $\{ p \}^{\{p, \neg p\}}$. Likewise, they must also have credence y in $\{ p \}^{\{p, \neg p\}}$. So by clause (B) of (4.32), $x = y$. ■

Theorem 4.35 If an agent α discerns better outcomes, their inquisitive credence state $\mathbf{QCr}_\alpha$ is an inquisitive probability.

Proof. If α does not discern better outcomes, they have no inquisitive credences (4.32). If they do discern better outcomes, their credences determine a function $\mathbf{QCr}_\alpha$ (clause (B) of 4.32). This function is defined on the space $\mathcal{Q}(\mathcal{D})$ of all the quizpositions about some domain of questions $\mathcal{D}$ (clause (D) of 4.32). The set $\mathcal{D}$ is closed under parthood (corollary 4.33).

This function $\mathbf{QCr}_\alpha$ satisfies the *Normality* condition of the definition (3.6) of an inquisitive credence: for any $Q \in \mathcal{D}$, $\mathbf{QCr}_\alpha(Q^Q) = 1$ (clause (A) of 4.32). It also satisfies the *Additivity* condition. For suppose Q contains R, A^Q and B^R are inconsistent. Then

$$\begin{aligned} \mathbf{QCr}_\alpha((A \cup QB)^Q) &= \mathbf{QCr}_\alpha(A^Q) + \mathbf{QCr}_\alpha(QB^Q) && \text{(by clause (C) of 4.32)} \\ &= \mathbf{QCr}_\alpha(A^Q) + \mathbf{QCr}_\alpha(B^R) && \text{(by lemma 4.35)} \end{aligned}$$

■

Theorem 4.18 An agent is inquisitive if and only if they discern better outcomes.

Proof. If the agent fails to discern better outcomes, they have no inquisitive credences or beliefs (4.32). If an agent does discern better outcomes, their credence state $\mathbf{QCr}_\alpha$ is an inquisitive probability. Thus they maximise utility (clause (E) of 4.32). And since inquisitive belief (4.13) is the same as inquisitive credence 1 (4.15), this entails that α's belief state $\mathbf{QB}_\alpha$ is an inquisitive information state (see footnote 23 above on p. 123). ■

Theorem 4.20 An agent maximises inquisitive expected utility when faced with Q if and only if they discern dominant Q-options.

Proof. If the agent does not discern dominant Q-options, by 4.32 they have no credences about

Q, and so there is no expected utility to maximise. If the agent does discern dominant Q-options, by part (E) of 4.32 they maximise expected utility when faced with Q. ■

Theorem 4.37 Suppose an agent α inquisitively believes the quizposition W^W, the discrete question $\{ \{w\} : w \in \mathscr{W} \}$. Then α is a classical agent.

Proof. By lemma 4.20, if α inquisitively believes W^W, then α discerns dominant W-options. But this is the same as discerning dominant options (see dfs. 4.11 and 4.19). So by theorem 4.12, α is classical. ■

Chapter 5.
Mathematical Belief

The quizpositional account of belief content is a hyperintensional theory: it makes distinctions between truth-conditionally equivalent belief contents. As explained in chapters 1-3, this feature resolves various problems for the classical view of belief and action. However, some other problems associated with intensional views of belief content are not directly resolved by the view. In this chapter, I will examine one such problem in detail, namely the problem of mathematical omniscience. This is the problem of how to understand beliefs about mathematics, and especially ignorance and uncertainty about mathematical issues. I will also make some remarks about another problem related to hyperintensionality, namely Frege's Puzzle. In both cases, I will argue that while the quizpositional account of belief does not by itself *solve* the issue, it does seem to bring us a step closer to a solution.

In the classical case, the problem of mathematical omniscience is divided into two parts, which might be called the problem of *explicit* mathematical omniscience and the problem of *implicit* mathematical omniscience. In §5.1, I explain the problem of explicit mathematical omniscience, and show why it initially looks like the move from intensional propositions to quizpositions makes no progress on this issue at all. In §5.2-3 I discuss the problem of *implicit* mathematical omniscience, and show that the quizpositional view does appear to solve that issue.

In §5.4, I return to the explicit problem and argue that a satisfactory solution to the explicit omniscience problem should allow us to make sense of uncertainty and credences about purely mathematical claims. I conclude that in order to achieve this within the constraints of the present account of belief, we will have to find a way of associating mathematical claims with contingent quizpositions. In §5.5-6, I examine two possible strategies for executing that project.

In §5.5, I consider the metalinguistic strategy advocated by Robert Stalnaker. First we consider this strategy in the context of the classical view of belief, and raise some important objections to this strategy due to Larry Powers and Hartry Field. While these objections seem devastating in the classical context, they do not carry over to the metalinguistic strategy as adapted to the inquisitive theory. §5.6 examines a different, metaphysical strategy for understanding mathematical ignorance, namely to endorse the unorthodox position that purely mathematical states of affairs are contingent. Again, we find that while this move does not bring us closer to resolving the problem on classical assumptions, it does have promise in the inquisitive context.

In §5.7, I turn to an aspect of Frege's Puzzle, namely the problem about informative identities (this is the issue that originally occupied Frege in the *Begriffsschrift*). Again, I consider a metalinguistic strategy for addressing this problem, which is essentially the suggestion Frege initially made. It turns out that many of the observations made about the metalinguistic strategy for addressing the problem of mathematical omniscience carry over to this context. For analogous reasons, the metalinguistic strategy fails in the context of the classical theory of belief, but not in the context of the inquisitive theory.

5.1 The Problem of Explicit Mathematical Omniscience

According to philosophical orthodoxy, the truths of pure mathematics are necessary truths, while pure mathematical falsehoods are always necessary falsehoods — call this metaphysical doctrine *mathematical fatalism*. Mathematical fatalism immediately creates an issue for the intensional account of propositions. For if it is true, then there are only two intensional mathematical propositions: the necessary truth $\top$ (the set of all possible worlds), and the necessary falsehood $\bot$ (the empty set). To make matters worse, the consistency requirement of the classical theory entails only the former proposition can be believed. Thus, on the classical account, everyone believes every mathematical truth with certainty. Or in other words, everyone is mathematically omniscient. On the other hand it is impossible to believe, or even lend credence to, any mathematical falsehood.

Unlike the intensional view, the quizpositional (question-directed) view of propositions distinguishes a variety of necessary truths: every question Q has its own tautology Q^Q. Furthermore, the quizpositional view of belief states allows for agents who believe some necessary truths but not others. So at first sight, it looks like the quizpositional view may have some additional resources for representing different states of mathematical knowledge.

However, when it comes to explicit, pure mathematical beliefs, it is not clear that the plurality of tautologies really makes any difference. The issue is that, given mathematical fatalism, there is still only one mathematical question. Take for instance the question *How many real roots does the quintic* $x^5 - 5 \cdot x^2 + 1$ *have?* Intuitively, this question has at least five possible answers. But since we are identifying questions with partitions of logical space, four of those answers correspond to the empty set, while the fifth contains all possible worlds. Hence this question, and every other purely mathematical question, imposes the trivial partition $\{\mathscr{W}\}$. So it would appear that the inquisitive view does not make progress over the classical view after all. There is still just one mathematical truth, $\{\mathscr{W}\}^{\{\mathscr{W}\}}$, which is known by everyone; and there is one mathematical falsehood, $\bot^{\{\mathscr{W}\}}$, that nobody believes. So on the face of it at least, the quizpositional view is still stuck with the bizarre conclusion that we are all mathematically omniscient.

This problem seems to show that both the intensional and quizpositional accounts of belief impose an unacceptable level of idealisation on believers. As Stalnaker once wrote, "One cannot treat a mathematician's failure to see all the deductive relationships amongst the propositions that interest him [as deviations from the norm of rationality] without setting aside all of mathematical inquiry as a deviation from rationality. But this would be absurd. Mathematical inquiry is a paradigm of rational activity, and a theory of rationality which excluded it from consideration would have no plausibility." (Stalnaker 1976, p. 87; cf. Chapter 6 below).

At the same time, it would not be quite fair, I think, to regard it as a very serious objection against the inquisitive account of belief in particular that it fails to tackle this issue right off the bat. I never claimed that quizpositions would solve every problem, and this particular problem touches on many difficult issues about the nature of mathematics outside of the scope of the present inquiry. It is entwined with a range of complex questions about the nature and content of pure mathematical beliefs that continue to present a mystery to everyone. For instance, the basic puzzles in Benacerraf 1965, 1973 are still widely regarded as unresolved.

Stalnaker (1984, 1999b) argued that the intensional account of belief simply brings out more starkly a problem about mathematical knowledge that everybody has to confront in one form or other. I am inclined to think that he was right about this. It is not as if advocates of syntactically structured propositions are in possession of a satisfactory, fully developed epistemology for mathematics. Witness for instance Haim Gaifman's (2004) difficulties in defining a coherent concept of credence that is applicable to arithmetical statements. There is a reason that large swathes of work in epistemology implicitly or explicitly brackets the case of mathematical knowledge and belief. Independently of the conception of cognitive content assumed, mathematical beliefs are poorly understood.

So it is not clear that the inquisitive/quizpositional theory of belief is really any worse off than other theories of belief with regard to this particular problem. But I think we can do better than that, and show that it is a good deal better off than the classical theory at least. The considerations set out below persuade me that the inquisitive turn is likely to be one important component of the solution to the tangled cluster of issues that converge in the problem of mathematical omniscience, although I confess I do not really know what that ultimate solution will look like. If I am right about that, it is surely an additional point in favour of the inquisitive theory.

5.2 The Problem of Implicit Mathematical Omniscience

In 2003, mathematician Timothy Pennings performed an experiment on his dog, a Welsh corgi named Elvis. He would take Elvis for walks around Lake Michigan and throw a ball into the lake. Elvis' task was to retrieve the ball as quickly as possible. This confronted Elvis with the decision problem represented in figure 9: at which point to jump into the water? Determining the optimal jumping point given the initial conditions is a non-trivial mathematical problem.

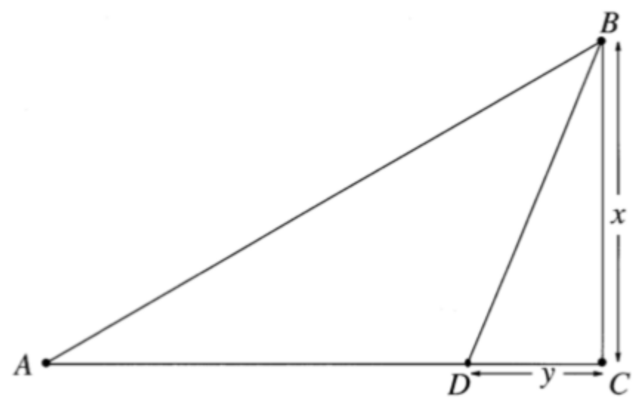


FIGURE 9: PATHS TO THE BALL (SOURCE: PENNINGS 2003)

In Figure 9, point *B* represents the spot where the ball lands, Elvis's starting point is the point *A*, and the line *AC* represents the edge of the water. Elvis is much faster running on land than swimming. In order to minimise his travel time, Elvis needs to jump into the water at just the right point *D* between *A* and *C*.

In Pennings' trials, he found that Elvis would identify the optimal jumping point with remarkable accuracy. He reported his findings in the *College Journal of Mathematics* under the title “Do Dogs Know Calculus?” (Pennings 2003). The editors of the journal were rather more willing to jump to conclusions, and ran with the headline “Elvis, the King of Calculus.” The article sparked a vivid debate about the nature and the extent of Elvis’ apparent mathematical knowledge (e.g. Dickey 2006, Perruchet and Gallego 2006, Bolt and Isaksen 2010).

It is not plausible that Elvis’ understanding of the situation amounts to explicit knowledge of the relevant mathematical facts. Pennings notes, for instance, that Elvis “has trouble differentiating even simple polynomials” (p. 182). Nonetheless, it is very natural to describe Elvis as exhibiting a kind of implicit mathematical knowledge: after all, he behaves as though he knows the solution to this abstract optimisation problem. For present purposes, I want to take this notion of implicit mathematical knowledge at face value, without attempting to analyse it.

Just as Elvis’ behaviour intuitively bears out some implicit mathematical knowledge, other behaviours bear out a lack of implicit mathematical knowledge. A particularly well-known example of the phenomenon, in humans, is the bat-and-ball puzzle from Shane Frederick and Daniel Kahneman (2002):

> A bat and a ball cost $1.10 together. The bat costs $1.00 more than the ball. How much does the ball cost?

In attempting this problem, almost everyone considers the answer “10 cents.” Most people stick with that initial hunch (Frederick 2005), but it is not the right answer. If the ball were to cost 10 cents, the bat would cost $1.10 *on its own*, and so the bat and ball would be $1.20 together, rather than $1.10. Once you see that 10 cents is too much, it becomes easy to guess and verify the correct answer: the ball costs 5 cents and the bat is $1.05.

There is a simple linear equation behind this exercise: $(1 + b) + b = 1.1$, where b is the cost of the ball. Those who derive the right answer can be said to have implicitly solved the equation: they exhibit a certain implicit mathematical knowledge. By the same token, the majority, who do not produce the right answer, implicitly fail to solve the equation correctly: you might say they exhibit a certain implicit mathematical ignorance.

Now one of Kahneman and Frederick's points with this this example is that the classical picture cannot account for what is going on. Frederick's respondents were some of America's finest young minds, all undergraduates at top institutions. So it is safe to assume that they fully understood the exercise. In particular, they know that the bat cost $1 + b$, and that the total price of bat *plus* ball is $1.10. Now those two pieces of information by themselves entail that the ball costs 5¢. So if the students had classical belief states, they would believe that the ball costs 5¢, since their beliefs would be closed under entailment. However, it is evident from their responses that the majority did not form that belief. The moral of the story is that the classical account does not only rule out the possibility of explicit mathematical ignorance, but also the possibility of implicit mathematical ignorance.

Here is a different, more interesting example of mathematical ignorance. Figure 10 below displays the ancient *Two Lovers* puzzle (also known as the Five Yen puzzle or Solomon's Seal). The initial setup is displayed on the left. The aim is to manoeuvre the two rings until they are adjacent to one another on the string, as in the diagram on the right. You cannot cut the string or break the board, and the rings are too large to pass through any of the holes.

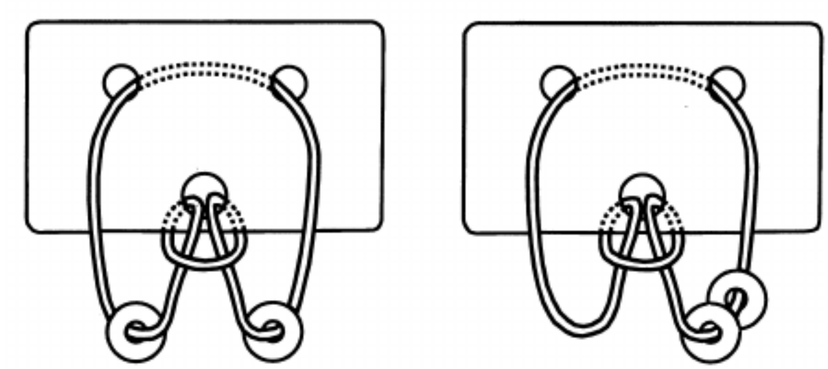


FIGURE 10: THE TWO LOVERS PUZZLE — HOW TO BRING THE RINGS TOGETHER? (SOURCE: HISAYOSHI 2003)

Although this can in fact be done, it is surprisingly difficult to find the solution. Puzzles like these can be viewed as physical instantiations of problems in topology or knot theory, and every step of the solution corresponds systematically to a step in the solution of the analogous maths problem (Nishiyama 2011; also Kauffman 1996, Horak 2006). In solving this puzzle, one is implicitly solving an abstract topological problem, and failure to see the solution is an implicit failure to solve to the isomorphic mathematical problem.

Again, the classical picture cannot account for people's behaviour in these cases. After a few minutes with this puzzle, anyone knows the basic mechanics of the construction. But these physical facts about the setup entail a method to arrive at a solution: at every possible world where the puzzle has thus and such physical structure, a certain particular sequence of steps leads to a solution. So on the classical account, anyone familiar with the puzzle should know how to solve it. However, it is evident from people's behaviour that this is not true.

Obviously this generalises. Whenever we are faced with a problem that can in principle be solved through mathematical calculation, it must be that the information we are given is sufficient to entail the solution. Thus the classical picture, according to which our beliefs are closed under entailment, predicts that in such cases one knows the solution in virtue of knowing the information we were given to start with. Implicit mathematical ignorance is manifested precisely when the believer in question fails to see the solution in spite of having all the necessary information. The classical picture does not allow for this possibility, whence it falsely predicts that implicit mathematical ignorance is impossible: this the problem of implicit mathematical omniscience.

(As the example of Elvis showed, implicit mathematical knowledge does not entail explicit mathematical knowledge. It might be worth noting that implicit mathematical ignorance, as roughly characterised by means of these examples, likewise does not entail explicit mathematical ignorance. A mathematician who knows the solution to the right abstract topological problem could in principle use this knowledge to solve the Two Lovers puzzle. But they are only in a position to do that once they see that the puzzle is a concrete instantiation of the relevant mathematical model. If the mathematician fails to make this link, they may still fail to solve the puzzle. Similarly, it is possible for someone who has just solved the equation $(1 + b) + b = 1.1$ to get tripped up by Kahneman and Frederick's question anyway, if they fail to make the connection between the equation and the story about the bat and the ball. In saying that agents' behaviours in theses cases manifest implicit ignorance of the mathematics, I just mean to say that the agent is manifestly failing to apply any knowledge of the relevant mathematics to the situation: this can happen whether the agent has the explicit mathematical knowledge in question or not.)

5.3 A Solution to the Implicit Problem

The advantage of focussing on the problem of implicit mathematical omniscience is that in doing so we can, for the moment, bracket all the deep metaphysical questions about the nature of mathematical propositions. The reason is that, in the cases of mathematical knowledge and ignorance described in the last section, there are no mysteries about the nature of the relevant explicit beliefs, which are about concrete, contingent matters. The relevant beliefs of Elvis the corgi, for instance, are beliefs about the position of the ball, running speed, where to jump to get to the ball quickly, etcetera. So for the purpose of explaining Elvis' behaviour, enigmas about the nature of mathematics can be left to one side.

Now what we find when we look at all these cases from an inquisitive perspective is that the implicit problem of mathematical omniscience has essentially dissolved. Take the bat-and-ball example for instance. We can assume that after reading the exercise, Frederick's respondents have gathered the quizpositional beliefs A^Q and B^R, where

- Q: *How much did the bat and ball cost together?*, A: *$1.10*
- R: *What is the difference between the price of the bat and the price of the ball?*, B: *$1*

And given their response, most respondents apparently drew the conclusion C^S:

- S: *What is the price of the ball?*, C: *$0.10*

Since the questions Q, R and S do not overlap, these three beliefs, while inconsistent, all fit together into a coherent inquisitive information state. And attributing this combination of beliefs to Frederick's respondents accounts for all their behaviour.

Those who did get the right answer, by contrast, ended up drawing the correct conclusion from A^Q and B^R, namely D^S:

- S: What is the price of the ball?, D: $0.05

Thus the inquisitive theory distinguish two different doxastic states, each of which explains the behaviour of the corresponding group. By contrast, the classical view can only explain the behaviour of those who got the right answer. In other words, the classical view assumes implicit as well as explicit mathematical omniscience, while the inquisitive view can easily accommodate this manifestation of mathematical ignorance.

Frederick 2005 provides two other nice examples, about which analogous points can be made. One is "It takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes." And "In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days." Most people instinctively answer "100" and "24" respectively. But some reflection reveals that the correct answers are in fact "5" and "47". All three puzzles cause trouble for the the classical view, while allowing for a straightforward quizpositional analyses.[30]

Likewise, in the Two Lovers case, we can represent the agent's knowledge of the situation by a collection of quizpositions about various aspects of the setup: *the thread goes this way, then through that hole, then that way*; *the rings are attached to the string on either side of the hole*; and so on. Different agents may vary as to which exact questions they take those beliefs to answer. But it is clear at the outset that none of those beliefs is going to have any immediate bearing on the questions they are faced when attempting to solve the puzzle, like *How do I get the ring to the other side?* or *What move should I make first?* A trial-and-error approach to the puzzle is precisely what you would expect, given this belief state. Thus the inquisitive account easily yields an explanation for people's typical response to this puzzle.

In principle, an alternative to trial-and-error is to try and reason your way to a solution. For your average person, this priori method is not likely to succeed on its own, because the issues at play are quite complex (Nishiyama 2011). But a knot theorist or topologist will have a better chance of success. Their mathematical experience gives them a better understanding of how to think about the situation, and of what transformations are possible. Or one might say that topologists have learnt what the right questions are to ask in this situation, and have

---

[29] These observations only establish that the inquisitive theory, unlike the classical theory, has the resources to represent the belief states of both the successful and the unsuccessful respondents. This is certainly progress, but it leaves other interesting questions in the neighbourhood unanswered: How did they acquire those belief states? What accounts for the difference between those who get the right answer and those who get the wrong answer? And especially: what explains the attractiveness of the illusory inference C^S ? These questions are beyond our present scope, but I do want to note the relevance of Philipp Koralus' and Salvador Mascarenhas' work in this context. These authors have shown how an inquisitive account of mental content can be used to account for illusory inferences and a wide range of common reasoning mistakes. See Koralus and Mascarenhas 2013, 2018, Mascarenhas and Koralus 2015, 2017, Parrott and Koralus 2015, Sablé-Meyer and Mascarenhas 2019.

internalised certain useful ways of dividing up logical space (this is in line with a suggestion by Pérez Carballo 2016 and the account of deduction provided in Chapter 2 above).

What about the genius corgi Elvis? Here the classical account gets the right result: assuming knowledge of the initial conditions, the classical theory correctly predicts that Elvis knows where to jump. But even in this case, I think the inquisitive account gives us a more realistic picture of what is going on. The inquisitive account admits the *possibility* of implicit mathematical knowledge: it is just not the only possibility. Presumably, Elvis has somehow learnt to link his observations about the ball to his opinion on where to jump. It seems unlikely that this mental process really involves the sorts of calculus that Pennings presents in his paper. More plausibly, the transition involves a quicker, more direct heuristic. In particular, it seems plausible that with practice, Elvis acquired a set of beliefs of the form *If the ball lands* ***there****, jumping* ***here*** *will get me to the ball the quickest*, which provide a direct link between his estimation of where the ball lands and his view on where he should jump.

The classical theory explains Elvis' success in terms of a necessary connection between his views on the initial conditions and his view of the jumping off point. The alternative explanation just proffered explains it in terms of Elvis' acquisition of a contingent and fallible heuristic that links the two. It strikes me that the latter explanation is preferable, being both more plausible and more empirically adequate. First of all, unlike the classical story, the inquisitive story can accommodate the fact that Elvis' judgment is less than perfectly accurate (Pennings' data show that on a few occasions, Elvis' estimates were far off the optimal mark). Secondly, because the possession of that heuristic is contingent, the inquisitive explanation does justice to the fact that

Elvis' knack for this is at least somewhat remarkable: there is a good reason that Elvis' picture was featured on the cover of the *College Journal of Mathematics*. The classical theory, by contrast, fails to account for the fact that many dogs would do a worse job at making the right call (as would most humans for that matter, even those that know a lot of calculus).

To sum up: even though the inquisitive account of belief does not come pre-packaged with a solution to the problem of explicit mathematical ignorance, it does yield a solution to the problem of implicit mathematical ignorance "straight out of the box," without any fancy footwork required. That should give us some courage as we investigate how the explicit problem of mathematical ignorance should be addressed.

5.4 Credence and Contingency

In the last section, I mentioned Alejandro Pérez Carballo's (2016) account of pure mathematical knowledge, according to which "to discover a new mathematical theory (or structure) is not to acquire a new [contingent] belief. Rather, it is to change the granularity of one's working picture of logical space—in other words, to change one's working hypothesis space." (p. 482) Pérez Carballo is rather vague about exactly what this comes to, but what he says is clearly reminiscent of the way that, in the present theory, believing a tautological quizposition Q^Q is not a matter of having information, but of partitioning the space of possibilities (see also Rayo 2013, and §2.5-10 above). So the most straightforward way of implementing Pérez Carballo's suggestion in the present framework would be to say that mathematical beliefs are beliefs in tautological quizpositions (that is, quizpositions of the form Q^Q).

To some extent, the discussion in the last two sections vindicates this idea: by learning new tautologies, one can come to see more consequences of ones beliefs, and thus gain implicit mathematical knowledge. Nevertheless, I do not think that this way of thinking about mathematical belief will ever yield a fully satisfactory account of mathematical knowledge and belief. One major issue is that it is unclear, given the issues pointed out in §5.1, how to associate mathematical claims with non-trivial questions. Pérez Carballo provides a suggestive example involving the bridges of Königsberg, but it is not at all clear how this is supposed to generalise even to other applications of graph theory.[30]

But even if there were a systematic way to map purely mathematical statements to tautological quizpositions, that still would not give us what we need. The reason is that, on the inquisitive theory of belief, one can only have two attitudes to a necessarily true quizposition: credence 1 or no attitude at all. Likewise, one can have two attitudes to a necessarily false quizposition: credence 0 or no attitude at all. But, as Haim Gaifman (2004) extensively argued, this is not at all satisfactory as an account of mathematical beliefs.

For one, it is possible to have false mathematical beliefs. To take an example from Gaifman,

[30] Elsewhere in the article, Pérez Carballo seems to be aware of this issue, and suggests that a mathematical belief is instead an *ability* to divide logical space in certain ways, where each individual division corresponds to an application of the mathematical theory (see e.g. bottom of p. 472). In the present case this would correspond to an *ability* to acquire certain beliefs of the form Q^Q. As far as I can see however, this more sophisticated formulation fails for somewhat analogous reasons to the ones I am about to adduce against the straightforward implementation. In particular, it is still entirely unclear, on Pérez Carballo's proposal, what it would be to have a *false* mathematical belief, or to be unsure about a particular mathematical proposition.

Marine Mersenne believed that $2^{67} - 1$ is prime, but in fact it is not. Or take the example of the Banach-Tarski theorem, which, according to Hartry Field (1978), would strike anyone unfamiliar with it as being highly implausible. And besides outright belief and disbelief, all intermediate doxastic attitudes seem possible and reasonable as well. For instance, it is the present consensus amongst number theorists that Goldbach's conjecture is extremely likely to be true. But most would stop short of saying they were absolutely certain of it. For example Neil Sheldon (2003) argues that your credence in the Goldbach conjecture should be $1 - 10^{-150,000,000,000}$. The ABC-conjecture, by contrast, is also considered quite probable, but less so.

Or take the twin prime conjecture, which says that there are infinitely many pairs of primes $(p, p + 2)$. It is not known whether this is true, but it is evidently more likely to be true than the stronger prime sextuplet conjecture, which says there are infinitely many sextuples of primes of the form $(p, p + 4, p + 6, p + 10, p + 12, p + 16)$ — the closest possible grouping of six primes. Intermediate between those two conjectures are the prime triplet, prime quadruplet and prime quintuplet conjectures — in that order.[31]

The only way to capture our graded doxastic attitudes about these prime tuples issue within the

[31] A prime triplet is a triplet of consecutive primes such that the first and the last are six apart — that is a triplet of the form $(p, p + 2, p + 6)$ or of the form $(p, p + 4, p + 6)$. Two overlapping triplets make a prime quadruplet $(p, p + 2, p + 6, p + 8)$. A prime quintuplet is a quadruplet with another nearby prime attached: $(p - 4, p, p + 2, p + 6, p + 8)$ or $(p, p + 2, p + 6, p + 8, p + 12)$, and a sextuplet is two overlapping quintuplets. The prime n-tuplet conjecture says there are infinitely many prime n-tuplets. None of these five conjectures have been proven or disproven, and no-one knows of a way to prove the stronger ones on the basis of the weaker ones — so they form a strictly increasing progression.

constraints of the present framework, is to treat them as contingent. What is needed is a way to separate the six intuitively distinct answers to this question P, *Which prime tuples are there infinitely many of*, so that we can represent it thus:

P = { *none of the tuples*; *only the twins*; *only the triples and the twins*;
the quadruplets, the triples and the twins; *all but the sextuplets*; *all of the tuples* }

If we could somehow manage to manoeuvre ourselves into this position, everything would work "as normal." We could distinguish thirty-two different quizpositions A^P about P in the ordinary way, and attribute agents a probabilistic view on P. It is easy to envisage how this approach would extend to other mathematical questions and beliefs. For instance, the Goldbach conjecture can be viewed as one answer to the question *Which even numbers cannot be written as the sum of two primes*?

So if only mathematical questions corresponded to non-trivial partitions, then we would be out of the woods. Pure mathematical claims would express ordinary quizpositions, and pure mathematical beliefs would be ordinary inquisitive beliefs. Thus we would avoid treating them as a fundamentally different kind of mental state from contingent beliefs about concrete matters. It would do justice to the intuition, which Gaifman emphasises, that mathematical inquiry is an ordinary part of the project of finding the truth, distinguished only by its generality (Gaifman 2004, p. 99).

If only. But the trouble is: where are we to find these non-trivial partitions? Consider again the question P, *Which prime tuples are there infinitely many of*. Assuming mathematical fatalism is true, there is just one possible state of affairs here, and that is all we have to work with. So all this speculation about what we could do if we could distinguish five possible state of affairs seems idle: the hard fact is that we can't. The same can be said of the question *Which even numbers are not the sum of two primes*, and the question considered in §5.1, *How many real roots does the quintic* $x^5 - 5 \cdot x^2 + 1$ *have*.

In the next two sections I will consider two approaches for overcoming this obstacle and associating mathematical claims with contingent propositions anyway. In §5.6, I consider the radical strategy of denying mathematical fatalism outright. But we will start with a more conservative approach, namely the "metalinguistic" strategy proposed by Robert Stalnaker.

5.5 The Metalinguistic Strategy

What is 73 times 100? Sophie, a bright seven-year old, knows the answer: it is 7300. What about 41 times 64? She could not tell you, or not without doing some calculation first. But what does her ignorance consist of in the latter case? What is it the calculation will reveal to her? Everyone can agree that amongst the things Sophie does not at this point know, and which she might come to know through performing the calculation, is which base-ten numeral co-refers with the expression "41 times 64".[33]

But is Sophie, in addition to being uncertain about that metalinguistic fact, also uncertain about the numbers themselves? That is to say, is her ignorance about the numerals a surface manifestation of an underlying ignorance about the relations that hold between the numbers themselves, an ignorance about which *number* the multiplication operator relates to the *numbers* 41 and 64? Likewise, does Sophie's knowledge that 73 · 100 equals 7300 amount to anything more than knowing that "73 · 100" co-refers with "7300"? That is, does she in addition possess some non-trivial piece of knowledge about the numbers these numerals represent?

Stalnaker (1976, 1984, 1990) argues that the answer to all of those questions is "no", and that belief reports on purely mathematical matters are systematically reinterpreted as beliefs about the mathematical symbols used in those reports. Stalnaker 1990 contains something like the following argument for the latter claim. It is clearly appropriate to describe Sophie's state of arithmetical knowledge as follows: "She knows that 73 times 100 is 1700, but she does not know what 41 times 64 is." But the following characterisation is inaccurate: "She knows that 111-base-eight times 124-base-eight is 1624-base-eight, but she does not know what 51-base-eight times 100-base-eight is!"

Insofar as we are concerned with Sophie's knowledge and ignorance about the numbers themselves, the latter descriptions should do fine. After all, 73 and 111-base-eight are both

---

[32] A nominalist may resist the idea that the expression "41 times 64" co-refers with some decimal because they deny that such expressions refer at all. But nominalists typically still admit a sort of faux coreference: a syntactically specifiable relation platonists misidentify as coreference, and which holds between "41 times 64" and a certain unique decimal. I will be sloppy on the distinction between faux coreference and coreference, and also about the distinction between faux mathematical truth and truth (where "faux mathematical truth" is some syntactically specifiable property like following-from-the-axioms). The defining feature of the metalinguistic view as I want to understand it here is this: the cognitive states that play the role of mathematical beliefs are beliefs about mathematical expressions, rather than being just about the mathematical objects and states of affairs that those expressions purport to refer to.

equally good names for the same number. But if we are re-interpreting these reports as claims about Sophie's knowledge of the relationship between certain representations, then it is clear why the latter report should strike us as inaccurate: for instance, it is simply not true that Sophie knows that "111-base-eight times 124-base-eight" corefers with "1624-base-eight".

There are also more theoretical reasons to think that mathematical ignorance is metalinguistic ignorance. One argument stems from the classical account of belief, which entails uncertainty about the necessary propositions $\top$ and $\bot$ is impossible. So if those propositions are the content of pure mathematical claims, the agent cannot be uncertain about the contents themselves. Another argument stems from a formalist conception of mathematics, according to which there just are no underlying mathematical states of affairs, and (much of) mathematical inquiry just is the study of the rule-bound manipulation of certain symbolic systems (e.g. Goodman and Quine 1947, Curry 1951, Stalnaker 1990, Weir 2015).

In his early work, Stalnaker fleshes out this view as follows: when ϕ is a purely mathematical sentence, belief attributions of the form 'α believes/knows that ϕ' should systematically be reinterpreted as 'α believes/knows that *the sentence ϕ is true*', (Stalnaker 1976; 1984, Ch. 4-5).[33],[34] This sort of view is also floated by Braddon-Mitchell and Jackson 2007, Ch. 11, and it is developed in Elga and Rayo 2018. Stalnaker (1976, 1984, 1990) claimed that this metalinguistic

[33] I am employing single quotation marks as Quine quotes here.

[34] Stalnaker 1990 retreats to a far less committal position on the contents of particular belief reports. However, as far as I can see, the points made below are largely independent of how exactly the metalinguistic strategy is implemented.

account of mathematical belief would allow him to distinguish different mathematical views while holding on to a coarse-grained, intensional conception of belief content.

Stalnaker's contention is based on a simple observation: besides the purely mathematical facts, the truth of the proposition that *the sentence "73 · 100 equals 7300" is true* depends on certain semantic facts: in particular, it depends on the details of the decimal number system. Even by the light of mathematical fatalism, those semantic facts are contingent. Now the truth of the propositions that *the sentence "111-base-eight times 124-base-eight equals 1624-base-eight" is true* and *the sentence "41 times 64 equals 2624" is true* each depend on quite different semantic facts, and thus these are all distinct propositions with different truth-conditions. Thus the intensional view of belief content is compatible with an agent believing any one of these metalinguistic propositions without believing the other two, even if, given the semantic facts as they actually happen to be, the three sentences that these three propositions attribute truth to happen to express the same proposition.

This is certainly an interesting and suggestive finding, but as Larry Powers (1976, p. 100), Hartry Field (1978, 1986, 2001) and Saul Kripke [35] were all quick to point out, it does not actually give us a solution to the problem of mathematical ignorance for the classical model of belief. The problem is the following. There are certain elementary facts that, by Stalnaker's own lights, are known by basically anyone with a primary school education, and which jointly entail the answer to every basic arithmetical question. So since, on the classical view, beliefs are closed under entailment, it still follows from this view that anyone who knows those elementary facts

---

[35] Kripke's point is reported in Stalnaker's *Inquiry*: Stalnaker 1984, p. 174, footnote 17 to chapter 4.

knows the answer to every basic arithmetical question. For this reason the classical picture still completely fails to capture the phenomenon of interest.

There are two ways to see this. The first argument runs as follows (Powers 1976, p. 100). Consider English translations of the axioms of Robinson arithmetic (Robinson 1950). These are all very elementary arithmetical claims, certainly within the scope of a Grade 2 education, so let's suppose Sophie knows that each one of these statements is true. Given how bright she is, it is not too much of a stretch to suppose Sophie also knows the axioms and inference rules of a simple system of first-order logic. Sophie knows, for instance, that *for any ϕ and ψ, if the sentence 'ϕ or ψ' is true but the sentence ϕ is not, then the sentence ψ must be true*. Now, the truth of the axioms, together with the validity of some inference rules, jointly entail the truth of every theorem of Robinson arithmetic.

So, if Sophie's beliefs really were closed under entailment, she would believe, of any theorem ϕ of Robinson arithmetic, *that ϕ is true*. In particular, this includes the solution of every basic arithmetical sum. If someone bothers to explain the second-order induction axiom to Sophie, she will be number-theoretically omniscient, and know the truth or falsehood of every single sentence in the language of arithmetic. So in spite of the metalinguistic twist, the classical view still dramatically fails to account for Sophie's mathematical ignorance.

The other path to this conclusion is even quicker (see Field 2001, p. 34-5, 101-2). Basic semantic competence with arithmetical terminology should entail that Sophie knows that "41" refers to the number 41, that "times" picks out the multiplication operation, that "64" refers to the number 64, that "equals" refers to the identity relation, and that "2624" refers to the number 2624. Since the underlying mathematical facts are necessary, those semantic facts alone entail that "41 times 64 equals 2624" is a true sentence. So because of deductive closure, it follows that Sophie knows that this is a true sentence, which evidently she does not.

Sure, the classical view allows for the possibility of agents who fail to know, of certain basic arithmetical truths, that they are truths. Cows for example, or mongooses — those agents lack knowledge of even the basic semantic facts involved. But the classical view does not allow for the possibility that English speakers with an elementary grasp of arithmetic, like Sophie, should fail to know that "41 times 64 equals 2624" is a true sentence. And it also does not allow for the possibility of a mathematician who knows that the axioms of a certain theory *T* are true, and that the inference rules are truth-preserving, but who fails to know of every theorem of *T* that it is true. So even with the metalinguistic move in place, the classical view still does not allow for any interesting cases of mathematical ignorance.[36]

Now note that these problems arise because the classical theory assumes closure under entailment. So it is worth investigating how the metalinguistic approach fares in the inquisitive

[36] Of course, Stalnaker is aware of these problems. His response has been to point to the possibility of belief fragmentation (Stalnaker 1984, p. 76-7, 1986, p. 120-3, 1999a; see also Elga and Rayo 2018). Given the commonalities between the inquisitive account and the fragmentation theory, the suggestion I am about to make is perhaps in the spirit to Stalnaker's broad-brushed remarks. However, as explained in §3.5 above, I do not view the inquisitive account of belief as a version of the fragmentation strategy. And I reject the notion that the ordinary failures of deductive closure should be understood as resulting from the fragmentation or compartmentalisation of our doxastic state, because I think that this view fatally undermines the essential unifying role that belief attributions play.

context. A metalinguistic analogue to the question like *What is 41 times 64?* is easily identified: *Which decimal numeral is coreferential with "41 times 64"?* Or alternatively *Which statement of the form "41 times 64 equals n" is true?* Or take the question considered in §5.4: *Which prime tuples are there infinitely many of*. A metalinguistic analogue is the question *Which of the prime tuple conjectures are true*, where these conjectures are understood as particular statements (in English or some formal language).

If we want a more systematic way of doing things, it is not difficult to specify general recipes for "lifting" a mathematical question Φ to a metalinguistic analogue. The details do not particularly matter for present purposes, but here is one approach:

1. Take any interrogative sentence Φ that expresses a mathematical question.
2. List statements that express distinct complete answers to Φ: $\phi_1, \phi_2, \ldots, \phi_k$
3. Let $\mathsf{m}_{\phi_i} = \{\, w : \text{at } w, \phi_i \text{ is true and } \phi_j \text{ is false for all } j \neq i \,\}$ and let $\mathsf{m}^* = \mathscr{W} \setminus \bigcup_i \mathsf{m}_{\phi_i}$
4. The "metalinguistic" content of Φ is the question M_Φ, *Which of* $\phi_1, \phi_2, \ldots$ *is true*:

$$\mathsf{M}_\Phi\text{: } \{\, \mathsf{m}_{\phi_1}, \mathsf{m}_{\phi_2}, \ldots\, \mathsf{m}_{\phi_k}, \mathsf{m}^* \}$$

Assuming fatalism is true, the semantic content of any complete answer ϕ to Φ is either $\{\, \mathscr{W} \,\}^{\{\mathscr{W}\}}$ or $\perp^{\{\mathscr{W}\}}$ — this what we saw in §5.1. However, the metalinguistic quizposition associated with ϕ is the quizposition $\{\, \mathsf{m}_\phi \,\}^{\mathsf{M}_\Phi}$. This is an ordinary, contingent quizposition: agents can be ignorant of them, they can reason about them, and they can accord them any level of credence.[37]

---

[37] What about incomplete answers ψ to Φ? Well, these express other quizpositions $\mathsf{A}_\psi{}^{\mathsf{M}_\Phi}$. It is natural to assume that any answer to M_Φ presupposes that the answers $\phi_1, \phi_2, \ldots, \phi_k$ are mutually exclusive and exhaustive, so that in general, $\mathsf{m}^* \notin \mathsf{A}_\psi$, and the agent's credence in $\{\, \mathsf{m}^* \,\}^{\mathsf{M}_\Phi}$ is equal to 0. But as long as $\varnothing \subset \mathsf{A}_\psi \subset \mathsf{M}_\Phi \setminus \{\mathsf{m}^*\}$, the quizposition $\mathsf{A}_\psi{}^{\mathsf{M}_\Phi}$ is just a run-of-the-mill, contingent quizposition.

Meanwhile, the objections to the classical version of the metalinguistic strategy do not go through here. Suppose Sophie is an inquisitive agent who believes that *whenever some sentence* ϕ *is true, so is '*ϕ *or* ψ*'*, and also that χ *is true* . It does not follow from these facts alone that Sophie knows that *'*χ *or* ψ*' is true*, even though that conclusion is entailed by her beliefs. In order to figure this out, Sophie will first have to reason about the sentences χ and 'χ *or* ψ' and put her beliefs together in the right way. Likewise, while Sophie's semantic beliefs might entail that *"41 times 64 equals 2624" is true*, it requires a non-trivial amount of deductive reasoning to see this consequence of her beliefs.[38]

Moreover, as a picture of pure mathematical belief and reasoning in particular, this language- and representation-centric picture of mathematical reasoning has a certain natural appeal. As Stalnaker puts it: "Mathematical information is most often received in linguistic form, and the behaviour that mathematical beliefs dispose us to engage in is primarily behaviour that involves linguistic and other representations: calculation, symbolic construction, and proof. It is not implausible, I think, to take representations and representational structures to be the subject matter of mathematics." (Stalnaker 1990, p. 237)

5.6 Rejecting Mathematical Fatalism

In §5.1, I premised the problem of mathematical ignorance on mathematical fatalism, the view that mathematical truths are necessarily true, and mathematical falsehoods are necessarily false.

[38] Note, incidentally, that we can also use this strategy to account for failures to recognise complicated logical truths as truths. By contrast, in the classical theory this metalinguistic way of accounting for ignorance about logic fails for the same reason as in the mathematical case.

This doctrine is a firmly established piece of philosophical orthodoxy: it is a highly intuitive view that continues to enjoy overwhelming support amongst philosophers. But there are in fact various good reasons to think that mathematical fatalism is false, which would open new perspectives on the problem of mathematical omniscience.

Call the negation of mathematical fatalism, the view that certain mathematical facts are contingent, *mathematical contingentism*. As I want to understand it here, this is the view that there are *possible* worlds where the mathematical facts are different. At those worlds, there might for instance be more or fewer mathematical objects, or the mathematical relations between those objects might be different in some way or other. I am *not* entertaining the suggestion that we enrich the logical space with impossible worlds.

Some arguments in favour of mathematical contingentism come from within mathematics itself. Certain mathematical results simply do not sit comfortably with the thesis that the mathematical realm is constant across all metaphysically possible worlds. In particular, it has been argued that phenomena related to "indefinite extensibility", such as the Burali-Forti paradox and the iterative conception of sets show that no matter how many mathematical objects there actually are, there could always be more of them (Parsons 1983, Hellman 1989, Hamkins 2012, Studd 2013, Linnebo 2013, 2018; McGee 2006 links related conclusions to the Löwenheim-Skolem theorem, Dummett 1963 and Wright and Shapiro 2006 to Gödel's Incompleteness Theorem.)

In addition, various general considerations from metaphysics and ontology favour mathematical contingentism. For instance, Gideon Rosen (2002) presents an argument that mathematical objects exist only contingently, based on the following two premises:

i) Mathematical objects are completely distinct from the concrete physical world (they share no common parts or constituents or ingredients with physical objects).

ii) Whenever two objects *A* and *B* are completely distinct, it is possible for an intrinsic duplicate of *A* to exist on its own (that is, without an intrinsic duplicate of *B* that is completely distinct from this duplicate of *A*).

Here (ii) is a general metaphysical principle advocated by, for instance, David Hume and David Lewis (Hume 1738, §1.3.6; Lewis 1986, §1.8). A number of other plausible general principles in ontology and metaphysical modality also tell against mathematical fatalism: see Field 1993, Colyvan 2000, Rosen 2002, Rayo 2013; compare Hale and Wright 1992, 1994. In addition to these general arguments, nominalists and fictionalists about mathematics have additional, strong reasons to take the non-existence of a mathematical objects to be a contingent matter (e.g. Burgess and Rosen 1997, Yablo 2005).

(I think a further consideration in support of contingentism is that we have a good error theory: there is a satisfying explanation for why mathematical fatalism would seem true even if it is in fact false. In other work, I have argued that statements that make reference to mathematical objects are routinely reinterpreted in ways that neutralise those ontological commitments. Now as it happens, given this same pragmatic mechanism, mathematical theorems express necessary truths even if they are literally false, and even if their literal content is a contingent proposition. This would explain why mathematical claims have the appearance of being necessary, even if the propositions they literally express are not in fact necessary; Hoek 2018, §4.7-5, see also Yablo 2014, 2017.)

Now these considerations in favour of mathematical contingentism offer some hope that the problem of mathematical ignorance might simply disappear. For if mathematical facts differ from one possible world to the next, then pure mathematical claims will be associated with contingent propositions, about which one can be ignorant or doubtful. However, those hopes are quickly dashed given the classical view of belief. For we still have closure under entailment to contend with. So it is still going to be true, on the classical account, that anyone who believes every axiom of a theory must believe every theorem of that theory as well.

For instance, suppose it turns out that the (finitely many) axioms of von Neumann–Bernays–Gödel (NBG) are all true, but only contingently so. Then it is still the case that *at every world where the axioms are true*, everything that follows from those axioms is also true. So on the classical account, everyone who believes each of the NBG axioms will believe every NBG theorem. And likewise, anyone who knows some very basic arithmetical truths knows the prime factors of 8,869,332,919. Thus the classical view falls prey to the problem of mathematical omniscience independently of whether mathematical fatalism is true or false.

But in the context of the inquisitive view, the contingentist strategy looks much more promising. Provided there is enough mathematical contingency, the inquisitive view allows for the possibility of an inquisitive agent who believes quizpositions corresponding to all the axioms of second-order PA, without believing, for instance, that there are infinitely many primes. After all, none of those axioms say anything about primes, or even use the concept. Thus none of the axioms group possible worlds according to the number of primes they contain. So while the axioms jointly *entail* that there are infinitely many primes, it is not *part* of any of the axioms, or even of the conjunction of all the axioms, that there are infinitely many primes.

The proviso that there be *enough* mathematical contingency is important. Mathematical contingentism could be true because only the continuum hypothesis and other undecidable statements of ZFC are contingent. If that were the case, it would not get us very far, and leave more pedestrian cases of mathematical ignorance, such as ignorance about the twin prime conjecture, unaccounted for. Or it might be that the existence of a Platonist realm is the only contingency. On this view, there are just two mathematical possibilities: it is possible that all mathematical objects exist and stand in the usual mathematical relations, and it is possible that there are no mathematical objects at all (Rosen 2002 considers this position). If we want to allow for interesting cases of mathematical ignorance, we will need more mathematical possibilities than that. However, it strikes me that it is at least prima facie plausible that there are a lot of alternative mathematical possibilities if there are any: once mathematical fatalism is given up, the floodgates are open so to speak, and it becomes difficult to see what principled limits on mathematical contingency might remain.

That is not to say that we need a contingentism so strong that almost *all* mathematical propositions are contingent: maybe the full solution to the problem of mathematical omniscience requires a hybrid contingentist-metalinguistic strategy. Perhaps in some cases, mathematical ignorance is due to ignorance about the mathematical facts, while in other cases it is just ignorance of semantic facts. It strikes me that some admixture of the metalinguistic strategy is intuitively mandated in any event. If you hear someone declare, with wide-eyed sincerity, that "One times one equals seven," the conclusion to draw is surely that the speaker does not know what those words mean.

I only provided some very bare-bones, programmatic sketches of how the metalinguistic and the contingentist strategies might look in the inquisitive context. Before we have a satisfactory basis for an epistemology of mathematics, a great deal more will need to be said, and further obstacles will have to be overcome. However, regardless of whether these particular strategies ultimately fail or succeed, the point stands that the inquisitive turn has improved their prospects substantially. In addition to this, it solves the implicit problem of mathematical omniscience. Taken together, I take this to be good reason that the inquisitive account brings us an important step closer to the solution of this problem, by giving us a principled way of rejecting closure under entailment.

5.7 Informative Identities

Intuitively, there is a difference between the belief that *Superman can fly* and the belief that *Clark Kent can fly*: Lois believes only the former. However, assuming the standard treatment of proper names as rigid designators (Kripke 1980), there is only one intensional proposition here. This is a well-known issue for the intensional view of belief, as well as many other views of belief content, including the Russellian view.

The move to quizpositions, as they are characterised in §1.3 above, does not immediately help. The polar question *Can Superman fly* partitions logical space into the same two parts as the question *Can Clark Kent fly* does. So while there is an intuitive difference between these two questions, our theory does not formally distinguish them. On the face of it, the inquisitive theory fails to distinguish the belief Lois has (*Superman can fly*) from the one she lacks (*Clark Kent can fly*). Similarly, the question *How far away is Hesperus* is the same partition as the question *How far away is Phosphorus*. But intuitively, an ancient astronomer who did not know that Hesperus is identical to Phosphorus may have different answers to these questions.

Relatedly, the reason that Lois's Superman-beliefs are separate from her Clark Kent-beliefs is that she falsely believes that *Superman and Clark Kent are different people*. But on the most straightforward treatment, all the questions that this belief could be taken to answer, like *Is Superman Clark Kent* or *Who is Superman*,induce the trivial partition $\{\mathscr{W}\}$. So Lois's belief, on this treatment, would seem to correspond to the trivial falsehood $\bot^{\{\mathscr{W}\}}$, which is not something one can believe. So while quizpositions are more finely individuated than intensional propositions, this additional fineness of grain does not immediately help address Frege's Puzzle. This problem about informative identities is clearly reminiscent of the situation with explicit mathematical ignorance.

As with the mathematical omniscience problem, I do not take this observation to count particularly heavily against the inquisitive theory of belief. After all, nobody has a fully satisfactory treatment of Frege's Puzzle. In particular, there are good reasons to think that a move to fine-grained, syntactically structured belief contents cannot solve Frege's Puzzle — see

Kripke 1979, Salmon 1983, Saul 1997, 2007, Dorr 2014, Goodman and Lederman 2018. But we can do a bit better, and in this section I describe how inquisitive decision theory may contribute to a solution to the issue about informative identities.

On the face of it, any true identity statement says of a certain object that it bears the identity relation to itself. But that seems like it is a triviality — so how can identity statements be informative? Frege's initial approach to this puzzle, in the *Begriffsschrift*, was metalinguistic (Frege 1879, §8). He surmised that in certain contexts, names like "Superman" and "Clark Kent" could refer to themselves instead of referring to their usual referent. The substantive fact that is expressed by the identity statement "Superman is Clark Kent", according to Frege, was that *"Superman" and "Clark Kent" are coreferential names*.

Although Frege (1892) retracted this view in favour of his theory of sense and reference, the metalinguistic approach survived. One modern incarnation of it is the diagonalisation theory devised by Robert Stalnaker (1978, 2004). According to this view, a statement ϕ can in certain conversational contexts express a different message from its literal content, which Stalnaker calls the *diagonal* proposition associated with ϕ. The nomenclature here is connected to Stalnaker's construction of this proposition in terms of two-dimensional modal logic. But there is no need to get into the weeds of that here. To a close enough approximation, the diagonal proposition associated with a sentence ϕ is just the intensional proposition that ϕ *is true* (Stalnaker 1987, p. 82). And the diagonal proposition associated with an identity claim 'α is β' is the proposition α *is coreferential with* β, exactly as Frege suggested a century earlier (ibid., p. 91; I ignore "Clintonite" cases where the meaning of "is" is in dispute).

In subsequent papers, Stalnaker (1981, 1987) extended this strategy to belief reports, claiming that reports of the form *α believes that ϕ* occasionally get reinterpreted as *α believes that ϕ is true*. Correspondingly, *Lois believes that Superman and Clark Kent are different people* gets to mean that *Lois believes that the person called "Superman" is different from the person called "Clark Kent"*.[39] For whatever reason, Stalnaker himself never makes the link, but it is clear that his metalinguistic approach to the problem of mathematical omniscience can seen as an instance of this type of embedded diagonalisation.

Thus it is perhaps no surprise that diagonalisation faces similar issues. The problem is that Lois has semantic information which entails that "Superman" and "Clark Kent" are coreferential. Lois knows who the man called "Superman" is. To say that she is acquainted with him would be an understatement: she has swayed in his strong arms as they floated through the cityscape at dusk, and shared a stolen kiss with him atop a skyscraper. And as all this transpired, Lois was constantly acutely aware that *this* man she was holding was none other than the person called "Superman". So she knows and believes that *"Superman" refers to this man*. She also knows all too well who the man called "Clark Kent" is: he is that pencil-sharpening, pitiful dork she sees slumping around the office every day — *that* guy. So clearly she knows and believes that *the person called "Clark Kent" is that man*. The trouble is that these two intensional propositions, *"Superman" refers to this man* and *"Clark Kent" refers to that man*, taken together, entail that *"Superman" and "Clark Kent" are coreferential*.

[39] In fact, Stalnaker also allows for further, half-diagonalised interpretations. In particular, sometimes a sentence of the form *α believes that β is γ* can get interpreted as *α believes that γ is the referent of 'β'*, and sometimes it can get interpreted as *α believes that β is the referent of 'γ'* (Stalnaker 1987, 127-9).

Sure, the classical view allows for agents who fail to believe that *"Superman" and "Clark Kent" are coreferential*. For instance, someone who has never even heard the names "Superman" or "Clark Kent" fails to believe this. (You and I also do not really believe it, because we think the name "Superman" does not refer at all.) But for the reasons I just explained, the classical account fails exactly in the interesting case: it cannot account for *Lois*'s ignorance, because her beliefs entail that *"Superman" and "Clark Kent" are coreferential*. Once again, this is strongly reminiscent of the mathematical case: the classical account allows agents that are unsure whether *Fermat's last theorem is true* — just no-one who knows basic arithmetic.

However, the example of Lois also brings out something more: the classical assumptions of closure under conjunction and entailment, when implemented possible-worlds style, imply idealisations that go beyond the attribution of an unrealistic level of computational power to agents. Presumably, a "computationally ideal" agent, whose mind is a perfect, instant Turing machine, could still be subject to identity confusions. So the rationality assumptions of the classical picture go beyond idealising agents' deductive abilities, introducing further distortions as well (§2.9 above made a similar point).

Once again, the inquisitive picture seems to avoid this trap. On the inquisitive view, the natural way to model Lois' background semantic beliefs is as follows: [40]

S: *To whom does "Superman" refer?*, A: *"Superman" refers to Kal-El*

K: *To whom does "Clark Kent" refer?*, B: *"Clark Kent" refers to Kal-El*

---

[40] Here I am following a convention to use Superman's Kryptonian name "Kal-El" as a more neutral alternative to "Clark Kent" and "Superman" — see for instance Pitt 2001.

Lois believes both A^S and B^K. In addition, she also believes C^R:

R: *Do "Superman" and "Clark Kent" refer to the same person or not?*, C: *They do not.*

While A^S, B^K and C^R are inconsistent, they are not incoherent, which is to say that they all fit together in an inquisitive belief state.

It is essential to this proposal that Lois does *not* believe the conjunction AB^{SK} of her semantic beliefs A^S and B^K. While the question R is part of neither S nor K, it *is* part of the question SK, so that $\neg C^R$ is part of AB^{SK}. That means AB^{SK} and C^R *are* incoherent, and cannot coexist in a single inquisitive belief state. The step from believing A^S and B^K to believing their conjunction AB^{SK} is one that Lois cannot make by deduction alone.

Some may regard this consequence of the view as unpalatable, but I take it to be independently motivated. Answering the questions S and K just requires Lois to keep track of the intrinsic semantic properties of the individual names, but answering the question SK in addition requires her to keep track of the semantic *relations* between "Superman" and "Clark", in particular the question of coreference. Thus the latter is a strictly more demanding task, which in this case requires Lois to sort out her identity confusions.

Kit Fine (2007) gave intriguing arguments to the effect that certain semantic facts about proper names are irreducibly relational. He analogised the situation to Russell's "antinomy of the variable" (Fine 2007, Ch. 1; Russell 1903, §3). Knowing the meaning of a variable in a first order language, "*x*" say, requires little more than knowing the domain over which "*x*" is allowed to range. But knowing the joint meaning of two variables "*x*" and "*y*" requires more than this: we need to know whether the variables are allowed to vary independently (as in standard predicate logic), or whether they are anti-correlated (as beginning students of predicate logic often assume). The inquisitive metalinguistic account of identity confusion just presented gives us a way to incorporate Fine's insight: it naturally renders knowing the reference of two names separately less demanding than knowing the reference of both names together. This gives us a clear way of understanding exactly how this issue about variables is supposed to be related to Frege's Puzzle (this is something that never became entirely clear from Fine's own exposition; see Pickel and Rabern 2017).

Bibliography

Akiyama, Hisayoshi, 2003, *Book of Entanglement Puzzles*. Tokyo: Shinkigensha.

Alchourrón, Carlos, Peter Gärdenfors, and David Makinson, 1985, "On the Logic of Theory Change", *Journal of Symbolic Logic* 50: 510–530.

Aloni, Maria, 2005 "Individual Concepts in Modal Predicate Logic." *Journal of Philosophical Logic* 34: 1–64

Armstrong, David, 1973, *Belief, Truth and Knowledge*. Cambridge University Press.

Arntzenius, Frank and Jeffrey Barrett 1999: 'An infinite decision puzzle'. *Theory and Decision* 46: 101-103.

Arntzenius, Frank, Adam Elga and John Hawthorne, 2004, "Bayesianism, Infinite Decisions and Binding." *Mind* 113(450): 251-283.

Aumann, Robert, 1987, "What is Game Theory Trying to Accomplish?". In *Frontiers of Economics*, ed. Kenneth Arrow and Seppo Honkapohja, 5-46. Oxford: Basil Blackwell.

Bacon, Andrew, 2018, "The Broadest Necessity." *Journal of Philosophical Logic* 47(5): 733-783.

Barcan Marcus, Ruth, 1983, "Rationality and Believing the Impossible." *Journal of Philosophy* 80(6): 321-338

Benacerraf, Paul, 1965. "What Numbers Could not Be." *Philosophical Review* 74 (1):47-73. (Repr. in Benacerraf and Putnam 1983, 272–294.)

1973. 'Mathematical Truth.' *Journal of Philosophy* 70(19): 661-679. (Repr. in Benacerraf and Putnam 1983, 403–420.)

Benacerraf, Paul and Hillary Putnam (eds.), 1983, *Philosophy of Mathematics: Selected Readings*. 2nd edition, Cambridge: Cambridge University Press,

van Benthem, Johan and Eric Pacuit, "Dynamic Logics of Evidence-Based Beliefs." *Studia Logica* 99: 61-92.

Berto, Francesco and Mark Jago, 2019, *Impossible Worlds*. New York: Oxford University Press.

Bjerring, Jens Christian, 2013, "Impossible Worlds and Logical Omniscience: An Impossibility Result." *Synthese* 190(13): 2505-2524.

Blaauw, Martijn, 2013, "Contrastive Belief." *Contrastivism in Philosophy: New Perspectives*, ed. Martijn Blaauw, 88-100. New York: Routledge.

Bledin, Justin and Kyle Rawlins, forthcoming, "Resistance and Resolution: Attentional Dynamics in Discourse." *Journal of Semantics*.

Bolker, Ethan, 1966, "Functions Resembling Quotients of Measures." *Transactions of the American Mathematical Society* 124: 293–312.

Bolt, Michael and Daniel Isaksen, 2010, "Dogs Don't Need Calculus." *College Mathematics Journal* 41(1): 10–16

Bradley, Richard. 1998. "A Representation Theorem for a Decision Theory with Conditionals." *Synthese* 116(2): 187-229.

Braithewaite, R.B., 1933, "The Nature of Believing." *Proc. of the Aristotelian Society* 33: 129 - 146.

Braddon-Mitchell, David and Frank Jackson, 2007, *The Philosophy of Mind and Cognition*. Second Edition, Oxford: Blackwell.

Bratman, Michael, 1984, "Two Faces of Intension." *Philosophical Review* 93(3): 375-405.

1987, *Intention, Plans and Practical Reasons*. Cambridge, MA: Harvard University Press.

Burgess, John and Gideon Rosen, 1997, *A Subject With No Object: Strategies for Nominalistic Interpretation of Mathematics*. New York: Oxford University Press.

Cherniak, Christopher, 1986, *Minimal Rationality*. Cambridge, MA: Bradford Books.

Ciardelli, Ivano, Jeroen Groenendijk, and Floris Roelofsen, 2013, "Inquisitive semantics: a new notion of meaning." *Language and Linguistics Compass* 7(9): 459-476.

Clarke, Roger, 2013, "Belief Is Credence One (in Context)." *Philosophers' Imprint* 13 (11): 1-18.

Colyvan, Mark, 2000, "Conceptual contingency and abstract existence." *Philosophical Quarterly* 50, 87–91.

Cresswell, Max, 1985, *Structured Meanings*. Cambridge, MA: MIT Press.

Curry, Haskell, 1951, *Outlines of Formalist Philosophy of Mathematics*. Amsterdam: North Holland.

Danto, Arthur, 1965, "Basic Actions." *American Philosophical Quarterly* 2(2): 141-148

Davidson, Donald, 1976, "Hempel on explaining Action" *Erkenntnis* 10(3): 239-253.

Davies, Martin and Andy Egan, 2013, "Delusion: Cognitive Approaches — Bayesian Inference and Compartmentalization." In: *The Oxford Handbook of Philosophy and Psychiatry*, eds. Fulford et al. Oxford: Oxford University Press, 689-727

Dennett, Daniel, 1971, "Intentional Systems." *Journal of Philosophy* 68: 87-106.

1975, "Brain Writing and Mind Reading." In: K. Gunderson (ed.), *Minnesota Studies in Philosophy of Science* VII. Minneapolis: University of Minnesota Press, 403-15.

Dickey, Leonid, 2006 "Do dogs know calculus of variations?" *College Mathematics Journal* 37(1):20–23.

van Ditmarsch, Hans, Wiebe van der Hoek, and Barteld Kooi, 2007, *Dynamic Epistemic Logic*. Dordrecht: Springer.

Dorr, Cian, 2014, "Transparency and the Context-Sensitivity of Attitude Reports." In: *Empty Representations*, Manuel García-Carpintero and Martí Genoveva (eds.), New York: Oxford University Press.

Dretske, Fred, 1970, "Epistemic Operators." *Journal of Philosophy* 67(24): 1007-1023.

1981, "The Pragmatic Dimension of Knowledge." *Philosophical Studies* 40(3): 363-378.

1988, *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Dummett, Michael, 1963. "The philosophical significance of Gödel's theorem." *Ratio* 5. Repr. in Dummett 1978, 186–214.

1975, "The Justification of Deduction." *Proceedings of the British Academy* 59: 201-232. Repr. in Dummett 1978, 290-318.

1978, *Truth and other enigmas*. Cambridge, MA: Harvard University Press.

Egan, Andy, 2008, "Seeing and believing: perception, belief formation and the divided mind." *Philosophical Studies* 140(1): 47–63

Egré, Paul and Denis Bonnay, 2012, "Metacognitive perspectives on unawareness and uncertainty." In: *Foundations of Metacognition*, eds. M.J. Beran et al., New York: Oxford University Press.

Elga, Adam, 2010, "Subjective Probabilities should be Sharp." *Philosophers' Imprint* 10.5, 1-11.

Elga, Adam, and Augustín Rayo, 2016, "Fragmented Decision Theory." Draft, dated Feb. 13.

2018, "Fragmentation, mathematical ignorance, and the metalinguistic reply." Draft, dated May 16. Posted online: jonathanweisberg.org/few2018/papers/elga_and_rayo.pdf

2019, "Fragmentation and Information Access." Draft, dated April 30. Posted online: www.princeton.edu/~adame/papers/fragment/fragmentation-and-information-access-2019-04-30.pdf.

Elliott, Edward, 2015, *Representation Theorems and the Grounds of Intentionality*. PhD thesis, Australian National University.

Fagin, Ronald, and Joseph Y. Halpern, 1988, "Belief, Awareness, and Limited Reasoning." *Artificial Intelligence* 34: 39–76.

Field, Hartry, 1978, "Mental Representation." Repr. in Field 2001, 30-67.

1986, "Stalnaker on Intentionality." *Pacific Philosophical Quarterly* 67(2): 98-112. (Repr. in Field 2001.)

1993, "The Conceptual Contingency of Mathematical Objects." *Mind* 102: 285-299.

2001, *Truth and the Absence of Fact*. Oxford: Clarendon Press 2001.

Fine, Kit, 2007, *Semantic Relationism*. Malden, MA: Blackwell.

2017, "A Theory of Truthmaker Content I, II" *J. of Philosophical Logic* 46(6): 625–702.

Fodor, Jerry, 1978, "Propositional Attitudes." *The Monist* 61: 501-23.

1980, "Methodological solipsism considered as a research strategy in cognitive psychology." *The Behavioral and Brain Sciences* 3: 63-109.

Foley, Richard, 1992, "The Epistemology of Belief and the Epistemology of Degrees of Belief." *American Philosophical Quarterly*, 29 (2): 111–24.

Franke, Michael and Tikitu de Jager, 2011 "Now that you mention it: Awareness Dynamics in Discourse and Decisions." In *Language, Games, and Evolution*, Benz et al. (eds), Lecture

Notes in Computer Science 6207, 60-91. Berlin: Springer.

Frederick, Shane, 2005, "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives,* 19(4), pp. 25-42.

Frege, Gottlob, 1879, *Begriffsschrift*. Halle: Verlag Louis Nebert.

1892, "Über Sinn und Bedeutung." *Zeitschrift für Philosophischen Kritik* 100: 25–50.

1884, *Die Grundlagen der Arithmetik*. Breslau: Verlag Wilhelm Koebner.

Friedman, Jane, 2017, "Inquiry and Belief." *Noûs* 53(2): 296-315.

2018, "Junk Beliefs and Interest-Driven Epistemology." *Philosophy and Phenomenological Research* 97(3): 568-583.

2019, "The Epistemic and the Zetetic." Unpublished manuscript.

Friston, Karl, and Christoph Mathys, 2017, "I am therefore I think." In: M. Leuzinger-Bohleber, S. Arnold and M. Solms (eds.), *The unconscious : a bridge between psychoanalysis and cognitive neuroscience*, 113-137. London: Routledge

Fritz, Peter, and Harvey Lederman, 2015, "Standard State-Space Models of Unawareness." *Proceedings of TARK XV*, Pittsburgh.

Gaifman, Haim, 2004, "Reasoning with Limited Resources and Assigning Probabilities to Arithmetical Statements." *Synthese* 140 (1-2): 97-119.

Gendler, Tamar, 2008, "Alief and Belief." *Journal of Philosophy* 105 (10): 634-663.

Gibbard, Alan, 2003, *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Gibbard, Alan, and William L. Harper, 1980, "Counterfactuals and Two Kinds of Expected Utility." In Harper, Stalnaker and Pears (eds.), 153-190.

Gilboa, Itzhak, 1987, "Expected Utility with Purely Subjective Non-Additive Probabilities," *Journal of Mathematical Economics* 16(1): 65-88.

Goodman, Jeremy, 2015, "Consequences of conditional excluded middle." Manuscript, dated February 25, 2015.

2018, "Agglomerative Algebras." *Journal of Philosophical Logic*, online first.

Goodman, Jeremy and Harvey Lederman, 2019, "Perspectivism." Draft dated 7 July, forthcoming in *Noûs*.

Goodman, Nelson and Quine, W. V., 1947, 'Steps towards a Constructive Nominalism', *Journal of Symbolic Logic* 12: 97–122.

Greco, Daniel, 2015a, "How I Learned to Stop Worrying and Love Probability 1." *Philosophical Perspectives* 29 (1):179-201.

2015b, "Iteration and Fragmentation." *Philosophy and Phenomenological Research* 91(3): 656-673.

Groenendijk, Jeroen, 1999, "The Logic of Interrogation." *Proc. of SALT* 9, 109-126.

Groenendijk, Jeroen and Martin Stokhhof, 1997, "Questions." In *Handbook of Logic and Language*, Johan van Benthem and Alice ter Meulen (eds.), 1055-1124. Amsterdam: Elsevier 1997.

Hacking, Ian, 1967, "Possibility." *Philosophical Review* 76(2): 143-168.

Hájek, Alan, 2005, "Scotching Dutch Books?" *Philosophical Perspectives* 19(1): 139-151.

Hale, Robert and Crispin Wright, 1992, "Nominalism and the Contingency of Abstract Objects." *Journal of Philosophy* 89: 111-135.

1994, "A *Reductio ad Surdum*? Field on the Contingency of Mathematical Objects." *Mind* 103: 169-184.

Hamblin, C.L., 1958, "Questions." *Australasian Journal of Philosophy* 36, 159-168.

Hamkins, Joel David, 2012 "The Set-Theoretic Multiverse." *Review of Symbolic Logic* 5(3): 416-449.

Harman, Gilbert, 1986. *Change in View.* Cambridge, Mass.: MIT Press

Harper, William, Robert Stalnaker and Glenn Pears (eds.), 1980, *Ifs*.. Dordrecht: D. Reidel.

Hart, J.T., 1965, "Memory and the feeling-of-knowing experience." *Journal of Educational Psychology* 56(4): 208–216.

Hawthorne, James, 2009, "The Lockean Thesis and the Logic of Belief." In: Franz Huber and Christoph Schmidt-Petri (eds.), *Degrees of Belief*, 49074. Synthese Library: Springer.

Hawthorne, John, Daniel Rothschild and Levi Spectre, 2016 "Belief is weak." *Philosophical Studies* 173(5): 1393-1404.

Hellman, Geoffrey, 1989, *Mathematics without Numbers: Towards a Modal-Structural Interpretation*. Oxford: Clarendon Press.

Hempel, Karl, 1962, "Rational Action." *Proc. of the American Philosophical Association* 35: 5-23.

Higginbotham, James, 2003, "Conditionals and Compositionality." *Philosophical Perspectives* 17(1): 181-194.

Hintikka, Jaakko, 1962, *Knowledge and Belief*. Ithaca: Cornell University Press.

1975, "Impossible possible worlds vindicated." *Journal of Philosophical Logic* 4(4): 475–484.

Hoek, Daniel, 2018, "Conversational Exculpature." *The Philosophical Review* 127(2): 151-196.

Holguín, Ben, 2019, "Believing the Answer to a Question." Draft, dated May 9.

Horak, Matthew, 2006, "Disentangling Topological Puzzles by Using Knot Theory". *Mathematics Magazine* 79(5): 368-375.

Hume, David, 1738, *Treatise of Human Nature*. London: John Noon.

Hutchinson, Pamela, 2010, "The world's hardest sudoku?" *The Guardian*, 22 August 2010.

Jago, Mark, 2013, "The Content of Deduction." *Journal of Philosophical Logic* 42(2): 317-334.

2015, "Hyperintensional Propositions." *Synthese* 192 (3): 585-601.

Jeffrey, Richard, 1974, "Savage's Omelette." *Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2, 361-371.

1982, "The Sure Thing Principle." *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 2, 719-730.

1990 [1965], *The Logic of Decision*. Second Edition, Chicago: University of Chicago Press.

Johnson-Laird, P.N., 1983, *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

2006, *How We Reason*. New York: Oxford University Press.

Joyce, James M., 1999, *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Kahneman, Daniel, 2012, *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, Daniel, and Shane Frederick, 2002, "Representativeness revisited: Attribute substitution in intuitive judgment." In: *Heuristics and biases: The psychology of intuitive judgment*, eds. Gilovich, Griffin and Kahneman, 49-81. Cambridge: Cambridge University Press.

Kahneman, Daniel, and Amos Tversky, 1979, "Prospect Theory: An Analysis of Decision under Risk." In *Econometrica* 47(2): 263-292.

Kant, Immanuel, 1785, *Grundlegung zur Metaphysik der Sitten*. Riga: Verlag J.F. Hartknoch.

Kauffman, Louis H., 1996, "Tangle complexity and the topology of the Chinese rings. In *Mathematical approaches to biomolecular structure and dynamics*, Mesirov et al. (eds.), IMA Volumes in Mathematics and its Applications 82, 1-10. New York, NY: Springer.

Kindermann, Dirk, Cristina Borgoni and Andrea Onofri (eds.), fc, *The Fragmented Mind*. Oxford University Press.

Klinedinst, Nathan, 2011, “Quantified conditionals and conditional excluded middle.” *Journal of Semantics* 28(1):149–170.

Koralus, Philipp, 2014, “The Erotetic Theory of Attention: Questions, Focus and Distraction.” *Mind & Language* 29.1: 26-50.

Koralus, Philipp, and Salvador Mascarenhas, 2013, “The Erotetic Theory of Reasoning” *Philosophical Perspectives* 27(1): 312-365.

2018, “Illusory inferences in a question-based theory of reasoning.” In: *Pragmatics, Truth, and Underspecification: Towards an Atlas of Meaning*, eds. Ken Turner and Laurence Horn, 300-322. Leiden: Brill.

Kratzer, Angelika, 2015 ”Chasing Hook.” Manuscript, dated May 2015. Forthcoming in: *Conditionals, probability, and paradox: themes from the philosophy of Dorothy Edgington*, Lee Walters and John Hawthorne (editors). New York: Oxford University Press.

Kripke, Saul, 1979, “A Puzzle about Belief.” In *Meaning and Use*, ed. A. Margalit, Reidel. 239-83.

1980, *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kyburg, Henry E. Jr., 1970. “Conjunctivitis.” In: Marshall Swain (ed.), *Induction, Acceptance, and Rational Belief*, 55-82. Dordrecht: D. Reidel.

Lavin, Douglas, 2013, “Must There Be Basic Action?” *Noûs* 47(2): 273-301.

Lehman, R.S., 1955, “On Confirmation and Rational Betting.” *Journal of Symbolic Logic* 20, 251–262.

Leitgeb, Hannes, 2014, “The Stability Theory of Belief.” *The Philosophical Review* 123(2): 131–171.

Levesque, Hector J., 1984, “A logic of implicit and explicit belief.” *AAAI-84 Proceedings*, 198-202.

1986, “Making believers out of computers.” *Artificial Intelligence* 30, 81–108.

Lewis, David, 1973, *Counterfactuals*. Oxford: Blackwell.

1980, “Counterfactuals and Comparative Possibility.” In Harper, Stalnaker and Pears (eds.), 57-85.

1981, “Causal Decision Theory.” *Australian Journal of Philosophy* 59(1): 5–30.

1982, “Logic for equivocators.” *Noûs* 16(3): 431-441.

1986, *On the Plurality of Worlds*. Oxford: Blackwell.

1988, “Relevant Implication.” *Theoria* 54(3): 161-74.

Linnebo, Øystein, 2013, “The Potential Hierarchy of Sets.” *The Review of Symbolic Logic* 6(2): 205-228.

2018, “Dummett on Indefinite Extensibility.” *Philosophical Issues* 28(1): 196-220.

MacKay, Alfred F. and Daniel D. Merrill (eds.), 1976, *Issues in the Philosophy of Language: Proc. of the 1972 Oberlin Colloquium in Philosophy*. New Haven: Yale University Press.

Maguire, E.A., D.G. Gadian, I.S. Johnsrude, C.D. Good, J. Ashburner, R.S. Frackowiak, C.D. Frith, 2000 “Navigation-related structural change in the hippocampi of taxi drivers.” *Proceedings of the National Academy of Sciences USA* 97: 4398-4403.

Maguire, E.A., K. Woollett and H.J. Spiers, 2006. “London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis.” *Hippocampus*, 16(12): 1091-1101.

Maguire, E.A. and K. Woollett, 2011, “Acquiring ‘the Knowledge’ of London’s Layout Drives Structural Brain Changes.” *Current Biology* 21(24): 2109-2114.

Makinson, David C., 1965, “Paradox of the Preface.” *Analysis* 25: 205-207.

Mascarenhas, Salvador and Philipp Koralus, 2015, “Illusory inferences: disjunctions, indefinites, and the erotetic theory of reasoning.” In: *Proc. of the 37th Meeting of the Cognitive Science Society*, eds. D.C. Noelle et al.

2017, “Illusory inferences with quantifiers.” *Thinking and Reasoning* 23(1): 33-48.

McGee, Vann, 1999, "An Airtight Dutch Book." *Analysis* 59, 257-265.

2006, "There is a Rule for Everything." In Rayo and Uzquiano 2006, 179-202.

Millikan, Ruth, 1986, "Thoughts Without Laws: Cognitive Science with Content." *The Philosophical Review* 95(1): 47-80.

Moss, Sarah, 2018, *Probabilistic Knowledge*. Oxford University Press.

Nagel, Thomas, 1971, "Brain Bisection and the Unity of Consciousness." *Synthese* 22: 396-413.

Nelson, D. L. and C. L. McEvoy, 1984, "Word Fragments as Retrieval Cues: Letter Generation or Search through Nonsemantic Memory?" *The American Journal of Psychology* 97(1): 17-36.

Neumann, John von and Oskar Morgenstern, 1953, *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Nishiyama, Yutaka, 2011, "Flexible geometry illustrated by the topology of an entangled ring puzzle." *International Journal of Pure and Applied Mathematics* 66(1): 61-69.

Norby, Aaron, 2014, "Against Fragmentation." *Thought* 3(1): 30-38.

Olofsson, U., L. Nyberg, 1995, "Determinants of word fragment completion." *Scandinavian Journal of Psychology* 36(1): 59-64.

Parikh, Rohit, 2009, "Sentences, belief and logical omniscience, or what does deduction tell us?" *Review of Symbolic Logic* 1(4): 459-476.

Parrott, Matthew and Philipp Koralus, "The erotetic theory of delusional thinking." *Cognitive Neuropsychiatry* 20(5): 398-415.

Parsons, Charles, 1983, "Sets and modality." In his *Mathematics in Philosophy*, 298–342. Cornell University Press, Cornell.

Peirce, Charles Sanders, 1878, "How to Make Our Ideas Clear." *Popular Science Monthly* 12: 286-302.

Pennings, Timothy J., 2003, “Do Dogs Know Calculus?” *College Mathematics Journal* 34(3): 178–182.

Pérez Carballo, Alejandro, 2016, “Structuring logical space.” *Philosophy and Phenomenological Research* 92(2): 460-491.

Perruchet, Pierre and Jorge Gallego, 2006, “Do dogs know related rates rather than optimization? *College Mathematics Journal* 37(1):16–18.

Phillips, L.D., 1966, “Some components of probabilistic inference.” *Technical Report* 1, Human Performance Center, University of Michigan.

Pickel, Bryan and Brian Rabern, 2017 “Does Semantic Relationism Solve Frege’s Puzzle?” *Journal of Philosophical Logic* 46: 97–118.

Pitt, David, 2001, “Alter Egos and Their Names.” *Journal of Philosophy* 98(10): 531-552.

Pollock, John L., 2002, “Rational choice and action omnipotence.” *The Philosophical Review* 111(1): 1-23.

Powers, Lawrence H., 1976, “Comments on Stalnaker, ‘Propositions’.” In MacKay and Merrill (eds.) 1976, 93-103.

1978, “Knowledge by Deduction.” *Philosophical Review*, 87(3): 337–371.

Priest, Graham, 2005, *Towards Non-Being*. Oxford: Clarendon Press.

Proust, Joëlle, 2013, *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. New York: Oxford University Press.

Ramsey, Frank, 1926, “Truth and Probability.” In Ramsey 1931, 138–154.

1931, *The Foundations of Mathematics and other Logical Essays*. Ed. R.B. Braithwaite, London: Kegan Paul.

Rayo, Agustín, 2013, *The Construction of Logical Space*. New York: Oxford University Press.

Rayo, Agustín and Gabriel Uzquiano, 2006, *Absolute Generality*. Oxford: Clarendon Press.

Raz, Joseph, 2011, *From Normativity to Responsibility*. New York: Oxford University Press.

Roberts, Craige, 2012, "Information Structure: Towards an integrated formal theory of pragmatics." *Semantics and Pragmatics* 5(6): 1-69.

Robinson, Raphael M., 1950 "An Essentially Undecidable Axiom System." *Proceedings of the International Congress of Mathematics 1950*, 729–730.

Rosen, Gideon, 2002, "A Study in Modal Deviance." In: *Conceivability and Possibility*, John Hawthorne and Tamar Gendler (eds.), 283-307. New York: Oxford University Press.

Ross, W.D., 1939. *Foundations of Ethics*. Oxford: Clarendon Press.

Rothe, Anselm, Brenden M. Lake and Todd M. Gureckis 2018, "Do people ask good questions?" *Computational Brain & Behavior*, 1, 69-89.

van Rooij, Robert, 2003, "Questioning to resolve decision problems." *Linguistics and Philosophy* 26: 727-763.

Russell, Bertrand, 1905, "On Denoting." *Mind* 14(56): 479–493.

Sablé-Meyer, Mathias and Salvador Mascarenhas, 2019, "Assessing the role of matching bias in reasoning with disjunctions." Draft, posted online: web-risc.ens.fr/~smascarenhas/docs/sable-meyer-mascarenhas19_matching_bias_disjunctions.pdf

Salmon, Nathan, 1983, *Frege's Puzzle*. Cambridge, MA: MIT Press.

Saul, Jennifer M., 1997, "Substitution and Simple Sentences." *Analysis* 57(2): 102–108.

2007, *Simple Sentences, Substitution and Intuitions*. New York: Oxford University Press.

Savage, Leonard J., 1967, "Difficulties in the Theory of Personal Probability." *Philosophy of Science* 34(4): 305-10.

1972 [1954], *The Foundations of Statistics*. Second Revised Edition. New York: Dover Publications.

Schaffer, Jonathan, 2009, “Contrastive Knowledge”. In: *Oxford Studies in Epistemology 1*, eds. Szabo Gendler and Hawthorne, 235-271. New York: Oxford University Press.

Schipper, Burkhard C., 2015, “Awareness.” In: *Handbook of Epistemic Logic*, eds. Hans van Ditmarsch et al., 77-146. London: College Publications.

Schmeidler, David, 1989, “Subjective Probability and Expected Utility without Additivity” *Econometrica* 57(3): 571-587

Schwitzgebel, Eric, 2002, “A phenomenal, dispositional account of belief.” *Noûs* 36: 249-275.

Shafir, Eldar, 1994, “Uncertainty and the difficulty of thinking through disjunctions.” *Cognition* 50: 403-430.

Shapiro, Stewart and Crispin Wright, 2006 “All Things Indefinitely Extensible.” In Rayo and Uzquiano 2006, 255-304.

Sheldon, Neil, 2003, “A Statistician’s Approach to Goldbach’s Conjecture.” *Teaching Statistics* 25(1): 12-13.

Sim, Kwang Mong, 2013, “Epistemic logic and logical omniscience: A survey.” *International Journal of Intelligent Systems* 12(1): 57-81.

Simons, Mandy, David Beaver, Craige Roberts and Judith Tonhauser, 2017. “The Best Question: Explaining the projection behavior of factive verbs.” In: *Discourse Processes* 54.

Singer, Murray, and Heather L. Tiede, 2008, “Feeling of knowing and duration of unsuccessful memory search.” *Memory & Cognition* 36, 588–597.

Skyrms, Brian, 1982, “Causal Decision Theory.” *Journal of Philosophy* 79(11): 695-711.

Smets, Sonja and Anthia Solaki, 2018, “The Effort of Reasoning: Modelling the Inference Steps of Boundedly Rational Agents.” In L.S. Moss et al. (eds.), *WoLLIC 2018*, Lecture Notes in Computer Science 10944, 307–324. Berlin: Springer.

Solaki, Anthi, Francesco Berto and Sonja Smets, 2019, “The Logic of Fast and Slow Thinking.” *Erkenntnis*. online first.

Stalnaker, Robert, 1968, “A Theory of Conditionals.” In *Studies in Logical Theory*, editor Nicholas Rescher, American Philosophical Quarterly Monograph Series 2, 98-112. Oxford: Blackwell.

1976, “Propositions.” In MacKay and Merrill (eds.) 1976, 197–214.

1978, “Assertion.” Repr. in Stalnaker 1999b, 225-240.

1980, “A Defense of Conditional Excluded Middle.” In Harper, Stalnaker and Pears (eds.), 87-104.

1981 “Indexical Belief.” Repr. in Stalnaker 1999b, 130-149

1984, *Inquiry*. Cambridge, MA: MIT Press.

1986, “Replies to Schiffer and Field.” *Pacific Philosophical Quarterly* 67(2): 113-123.

1987 “Semantics for Belief.” Repr. in Stalnaker 1999b,117-129

1990, “Mental Content and Linguistic Form.” Repr. in Stalnaker 1999b, 225-240.

1991, “The Problem of Logical Omniscience I.” Repr. in Stalnaker 1999b, 241-254.

1999a, “The Problem of Logical Omniscience II.” In Stalnaker 1999b, 255-273.

1999b, *Context and Content*. New York: Oxford University Press.

2004, “Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics.” *Philosophical Studies* 118(1-2): 299–322.

Storrs-Fox, David, 2019, “Non-Cartesian Agency and Normative Theory.” Draft, dated June 26, 2019. Posted online: sites.google.com/nyu.edu/davidstorrsfox/np-pdf.

Studd, James P., 2013, “The iterative conception of set: A (bi-)modal axiomatisation.” *Journal of Philosophical Logic* 42(5): 697–725.

Thaler, Richard, 2015, *Misbehaving: The Making of Behavioral Economics*. New York: W.W. Norton.

Thaler, Richard and Carl Sunstein, 2008 *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Toplak, Maggie and Keith Stanovich, 2002, “The Domain Specificity and Generality of Disjunctive Reasoning.” *Journal of Educational Psychology* 94(1): 197-209.

Tversky, Amos and Daniel Kahneman, 1990, “Advances in prospect theory: Cumulative representation of uncertainty.” *Journal of Risk and Uncertainty* 5(4): 297–323.

Tversky, Amos and Eldar Shafir, 1992. “The disjunction effect in choice under uncertainty.” *Psychological Science* 3: 305-309.

Vogel, Jonathan, 1990, “Are there Counterexamples to the Closure Principle?” In: *Doubting*, eds. Michael D. Roth and Glenn Ross, Philosophical Studies Series 48. Dordrecht: Springer.

Weatherson, Brian, “Games, Beliefs and Credences.” *Philosophy and Phenomenological Research*, 92(2): 209-236.

Weir, Alan, 2015, “Formalism in the Philosophy of Mathematics.” *The Stanford Encyclopedia of Philosophy*, Spring 2015 Edition, ed. Edward N. Zalta.

Williams, J. Robert, 2010 “Defending Conditional Excluded Middle.” *Noûs* 44(4): 650–668.

Williamson, Timothy, 2007, “‘How probable is an infinite sequence of heads?’ *Analysis* 67: 173-180.

2018, "Knowledge, Credence, and the Strength of Belief." Draft, dated April 18, posted online: media.philosophy.ox.ac.uk/docs/people/williamson/Strengthofbelief.pdf.

Yablo, Stephen, 2005, "The Myth of the Seven." In *Fictionalist Approaches to Metaphysics*, ed. Mark Eli Kalderon, 88-115. New York: Oxford University Press.

2006, "Non-Catastrophic Presupposition Failure." In *Content and Modality: Themes From the Philosophy of Robert Stalnaker*, eds. Judith Thomson and Alex Byrne, 164-190. New York: Oxford University Press.

2014, *Aboutness*. Princeton, NJ: Princeton University Press. With an online appendix at assets.press.princeton.edu/releases/m10013.pdf.

2017, "If-Thenism." *Australasian Philosophical Review* 1(2): 115-132.

Yalcin, Seth, 2007, "Epistemic Modals." *Mind* 116(464): 983-1026.

2008, "Modality and Inquiry." PhD thesis, Massachusetts Institute of Technology.

2011, "Nonfactualism About Epistemic Modality." In *Epistemic Modality*, eds. Andy Egan and Brain Weatherson, 295-332. New York: Oxford University Press.

2018, "Belief as Question-Sensitive." *Philosophy and Phenomenological Research*, 97(1): 23-47.